Combining library screening approaches, and modifying peptides with helix constraints, to generate novel antagonists of oncogenic Activator Protein-1

Baxter, Daniel

Award date:
2017

Awarding institution:
University of Bath

Combining library screening approaches, and modifying peptides with helix constraints, to generate novel antagonists of oncogenic Activator Protein-1

Daniel Baxter

A thesis submitted for the degree of Doctor of Philosophy

University of Bath
Department of Biology and Biochemistry

September 2016

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to express the deepest and sincerest gratitude to my family for all the support they have given me, consistently, throughout my PhD and previous studies, without which I would not be writing this manuscript.

I'd like to thank my supervisors Dr. Jody Mason, and Dr. Chris Ullman at Isogenica Ltd., for all the discussions, advice, support and encouragement these last five years that we've worked together.

I would also like to thank Dr. Laura Frigotto at Isogenica Ltd. for her invaluable discussions, encouragement and technical insights.

I want to thank my friends outside of the lab, and all my colleagues in the labs at the University of Bath and the University of Essex for being there to share some great times and some tougher ones, and for their advice, both technical and trivial. In particular, I would like to say a special thanks to Dr. Anna Panek and Dr. Victoria Allen-Baume. I'd similarly like to thank my Isogenica colleagues for their help, and their welcoming and friendly attitude towards me.

Finally, I'd like to express my gratitude to the BBSRC for funding this CASE PhD.

Dedication

This manuscript is dedicated to the memories of family members Meg Norman, Ruediger Kanne, Peter Whitehead and Dave Whitehead, who sadly passed away during my time in research.

Declaration of material from a previously submitted thesis/portfolio and of work done in conjunction with others

The following materials and data in Chapter 3 were produced by our collaborators Dr. Nathan Zaccai (NZ) and Prof. R. Leo Brady (RLB) at the University of Bristol, or Mr Samuel Perry (SP), Dr. Timothy A. Hill (TAH), Dr. W. Mei Kok (WMK) and Prof. David P. Fairlie (DPF) at the Institute for Molecular Biosciences, University of Queensland, Australia.

Peptides in Figure 3.2 were synthesized, purified and masses confirmed (Appendix Figure 2.6B-D and 2.7B-D) by TAH and WMK unless otherwise stated in section 2.9.2.

Figure 3.1 and Appendix Table 3.1: FosW–cJun crystal screens, diffraction data collection and model building were performed by NZ and RLB.

Figure 3.7 and Appendix Figure 3.2 FACS and microscopy data, and Figure 3.8 qRT-PCR assay data and associated statistical analyses, were performed by SP.

SP has submitted part of the above data, analyses and interpretation for the Degree of Doctorate in Philosophy at the University of Queensland, Australia, in August 2016.

All remaining data, and all text in this thesis, is the author's own work.

Publications and Presentations

Baxter, D., Ullman, C. U., Mason, J. M. (2014) Library construction, selection and modification strategies to generate therapeutic peptide-based modulators of protein-protein interactions, *Future Medicinal Chemistry*, 6(18), 2073-2092.

Crooks, R. O., **Baxter, D.**, Panek, A. S., Lubben, A. T., Mason, J. M. (2016) Deriving Heterospecific Self-assembling Protein-Protein Interactions Using a Computational Interactome Screen, *Journal of Molecular Biology*, 428, 385-398.

Oral Presentations

Gordon Research Conference "Peptides: Chemistry & Biology Of", Feb. 2016, Ventura; "Cell-Permeable Helix-Constrained Peptides Inhibit Activator Protein-1 and Reduce Cancer Cell Viability".

2 minute 'Lightning Talk', Graduate School Research Day, Jun. 2015, University of Bath; "Exploring Helicity Induction in Peptide Antagonists of the Activator Protein-1 Coiled Coil".

5th RSC Protein & Peptide Early Stage Researcher Meeting, Nov. 2014, RSC London; "Combining Library Selection and Modification Strategies to Generate Peptide-Based Antagonists of oncogenic Activator Protein-1".

Poster Presentations

Gordon Research Seminar "Peptides: Chemistry & Biology Of", Feb. 2016, Ventura.

41st Lorne Conference on Protein Structure & Function, Feb. 2016, Lorne.

11th Cancer Research at Bath Symposium, April 2015, University of Bath.

3rd PPI-Net Young Researcher's Symposium, Apr. 2014, Imperial College London.

Abstract

Activator Protein-1 (AP-1) is an oncogenic transcription factor that is dysregulated in numerous human cancers, making it an attractive therapeutic target. AP-1 forms via interaction of cJun and cFos proteins, which intertwine to generate a 'coiled coil' (CC) structure. Thus, the cJun/cFos α-helical CC domains responsible for dimerisation are appealing targets for inhibiting AP-1 formation and activity. Helical peptide antagonists that sequester cJun can be derived from the cFos CC domain by selection of more optimal amino acids for increased binding affinity. Peptides can then be downsized and modified to improve therapeutic potential. Two approaches aimed to identify novel short peptides against cJun. The first was to covalently cyclise amino acid side chains in existing cFos-derived peptide "FosW", with the aim of constraining FosW into a stable helix to allow downsizing without significant loss of binding structure and affinity. Using circular dichroism spectroscopy and isothermal titration calorimetry, a series of helix constrained peptides were characterised, from which a peptide was identified that retained 88 % of FosW binding affinity whilst being 22 % shorter, and which entered breast cancer cells *in vitro*, with preliminary data suggesting potential ability to inhibit AP-1 *in cellulo*. The second approach was to combine two existing high-throughput peptide selection systems, with the aim of benefitting from overlap in their strengths and weaknesses. Combination of *in vitro* CIS display and *in cellulo* Protein-fragment Complementation Assay successfully isolated a high affinity peptide from a hugely diverse library, and future refinements to further exploit this approach, particularly for short peptide selection, were formulated. Thus, molecules and techniques derived here may expedite the future development of therapies for cancers featuring AP-1 dysregulation.

List of Abbreviations

2XYT	2X Yeast Extract Tryptone medium
AP-1	Activator Protein-1
APS	ammonium persulphate
AU	Absorbance Units
bCIPA	bZIP Coiled-coil Interaction Prediction Algorithm
BOW	Base Optimised Weights
bp	base pairs
BSA	Bovine Serum Albumin
bZIP	basic Leucine Zipper
CANDI	Competitive And Negative Design Initiative
CC	Coiled Coil
CD	circular dichroism
cFos	cellular-Fos
cJun	cellular-Jun
CPP	Cell-Penetrating Peptide
CREB	cyclic AMP responsive element-binding protein
ddH_2O	double-distilled water
DHFR	Dihydrofolate Reductase

DMSO	dimethylsulfoxide
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FD	Fast Digest
FITC	Fluorescein Isothiocyanate
Fmoc	9-fluorenylmethoxycarbonyl
Fra-1/2	Fos-related antigen 1 or 2
FRET	Fluorescence Resonance Energy Transfer
F_U	Fraction unfolded
GdnHCl	Guanidinium hydrochloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-Performance Liquid Chromatography
IC_{50}	Inhibitor Concentration giving 50 % target activity reduction
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
ITT	*In vitro* transcription-translation
KAc	potassium acetate
KCl	potassium chloride

K_d	dissociation constant
KPP	potassium phosphate buffer
LB	Luria Bertani
LZ	Leucine Zipper
mDHFR1	murine dihydrofolate reductase fragment 1
mDHFR2	murine dihydrofolate reductase fragment 2
MEM	Modified Eagle's Medium
MES	2-(N-morpholino)ethanesulfonic acid
MMP	Matrix Metalloproteinase
MRE	Mean Residue Ellipticity
MS	Mass Spectrometry
MTT	3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide
NaAc	sodium acetate
NaOH	sodium hydroxide
NFAT	Nuclear Factor of Activated T cells
NLS	Nuclear Localisation Signal
NMR	Nuclear Magnetic Resonance spectroscopy
OD_{600}	Optical Density at 600 nm
PBS	Phosphate-Buffered Saline

PCA	Protein-fragment Complementation Assay
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank (RCSB)
PMA	phorbol 12-myristate 13-acetate; also known as TPA
PPI	Protein-Protein Interaction
RP-HPLC	Reverse Phase-High Performance Liquid Chromatography
RT-PCR	Reverse Transcription-PCR
SD	Standard Deviation
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEM	Standard Error of the Mean
SPPS	Solid Phase Peptide Synthesis
TAE	Tris-Acetate EDTA
TEMED	Tetramethylethylenediamine
T_m	melting temperature
TPA	12-O-tetradecanoylphorbol-13-acetate; also known as PMA
Tris	Tris (hydroxymethyl)-aminomethan

Chapter 1 – Introduction

1.1 Activator Protein-1 as a target for therapy of human cancers

Normal cellular health is dependent on carefully balanced responses to environmental signals. Within cells these are most commonly achieved via alteration of protein expression levels and patterns, such that the function and levels of activity of protein networks co-operating on specific tasks are adjusted accordingly. Protein expression is altered by nuclear transcription factors (TFs) that are activated or inhibited by transduced signals. Many human pathologies result from disruption of the balance between stimuli and responses that usually maintains homeostasis. Many of these, notably including cancers, can be attributed to transcription factor dysfunction/dysregulation, as is the case of oncogenic Activator Protein-1 (AP-1).

1.1.1 The normal role of Activator Protein-1

Activator Protein-1 (AP-1) is a mammalian dimeric transcription factor responsible for the transcriptional regulation of genes controlling cellular survival, proliferation, differentiation and apoptosis (Eferl and Wagner, 2003, Shaulian and Karin, 2002, Angel and Karin, 1991). AP-1 activity in actively proliferating cells is affected by survival- and proliferation-regulatory extracellular signals whose signalling pathways converge on AP-1. These pathways contribute to both normal cellular health and oncogenesis involving AP-1 (Figure 1.1) (Eferl and Wagner, 2003).

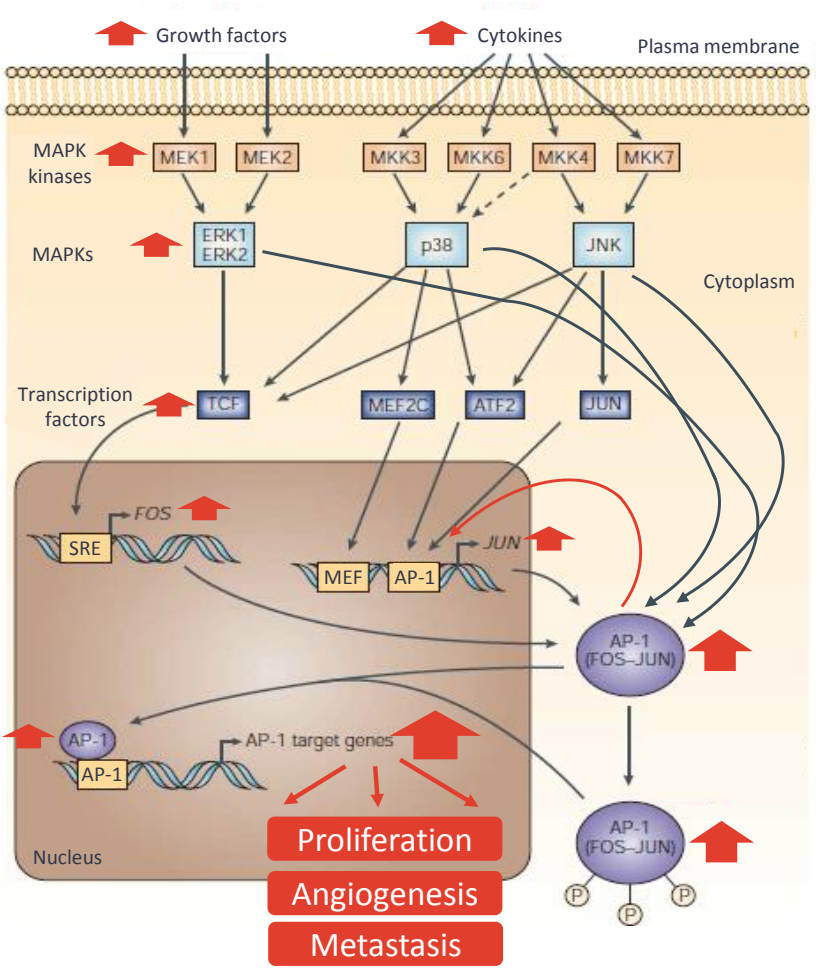


Figure 1.1: Pro-oncogenic growth factor and cytokine cellular signalling pathway activation of Activator Protein-1 (AP-1) to drive an oncogenic phenotype. Select activation pathways of AP-1 in response to environmental signals are depicted. Red block arrows denote high levels of protein expression or protein activity that overload normal activity of the pathways shown (grey arrows) to result in a gene expression pattern that drives an oncogenic phenotype of uncontrolled proliferation, abnormal angiogenesis, and metastasis. The auto-upregulation of cJun expression is highlighted with a thin red arrow. Figure adapted from source (Eferl and Wagner, 2003).

Extracellular stimuli affecting AP-1 activity include growth factors (GFs), cytokines, neurotransmitters, hormones, extracellular matrix components, and physical and chemical stresses (Shaulian and Karin, 2002). These alter expression of AP-1 component proteins of the Jun and Fos families, and activation of assembled

AP-1, via signalling pathways involving cell surface receptors and cytoplasmic signalling intermediaries, most notably including the mitogen-activated protein kinases (MAPKs). Signals are transduced via MAPKs such as Extracellular signal-Regulated Kinase (ERK) and Jun N-terminal kinase (JNK) to TFs including AP-1 (Eferl and Wagner, 2003). These TFs alter Fos and Jun expression via transcriptional control elements in *FOS* and *JUN* gene promoters (Angel and Karin, 1991).

Expressed Jun (cJun, JunB, and JunD) and Fos (cFos, FosB, Fra-1, and Fra-2) family proteins assemble via a protein-protein interaction (PPI) to form AP-1 (Eferl and Wagner, 2003). Functional AP-1 binds most commonly to target gene transcriptional control elements with the 12-*O*-tetradecanoylphorbol-13-acetate (TPA) responsive element (TRE) consensus sequence TGAg/cTCA. AP-1 can 'transactivate' transcription by RNA polymerase recruitment through interaction with TATA-binding protein and TFIIB, or via co-activators such as CREB-binding protein (Franklin *et al.*, 1995, Albanese *et al.*, 1999). Alternatively, transactivation incompetent AP-1 compositions such as cJun–JunB can inhibit expression (Deng and Karin, 1993), or AP-1 can repress transcription with the aid of co-repressors such as histone deacetylases for chromatin compaction (Bakin and Curran, 1999). Gene transcription activation or repression is dependent on AP-1 dimer composition, in turn dependent on AP-1 component and co-activator/co-repressor expression levels, which are cell- and cell-cycle stage-dependent (Eferl and Wagner, 2003, Miller *et al.*, 1984). Different AP-1 dimers bind to TRE sites, and transactivate transcription, more effectively than others (Ryseck and Bravo, 1991, Chiu *et al.*, 1989).

Target genes for AP-1 span proliferation/survival proteins (e.g. Cyclin D1 and p53), motility and invasion factors (e.g. matrix metalloproteinases (MMPs)), angiogenesis factors (e.g. vascular endothelial GF), and variety of other proteins with diverse functions (Eferl and Wagner, 2003). By controlling the expression of these genes, AP-1 contributes to diverse processes including foetal development and organogenesis, immune system function, wound healing and cognitive functions (Shaulian and Karin, 2002). However, AP-1 activation and activity can also contribute to tumourigenesis.

1.1.2 AP-1 dysregulation and cancer

Strict regulation of basal AP-1 activity in normal proliferating cells is crucial to maintain homeostasis. Complex transcriptional and posttranslational mechanisms regulate AP-1 activity (Angel and Karin, 1991). At the transcriptional level, these include auto-upregulation of cJun transcription, balanced by competition for the TRE site in the cJun promoter by transcriptionally inactive JunB- and JunD-containing AP-1, and auto-downregulation of cFos transcription. At the posttranslational level, JNK and ERK phosphorylation of cJun and cFos enhances AP-1 transactivation capability (Smeal *et al.*, 1991) through recruitment of co-activators like CREB-binding protein (Arias *et al.*, 1994), and can inactivate proteasome degradation signalling motifs within Fos and Jun proteins (Ferrara *et al.*, 2003).

In pro-oncogenic situations, an overabundance of mitogens, and hyperactive or constitutive MAPK signalling can overload AP-1 regulatory mechanisms (Figure 1.1). Dysregulated AP-1 component expression and posttranslational activation drives expression of a tumourigenic gene expression profile, featuring upregulation

of pro-proliferative, pro-angiogenic and pro-motility proteins and downregulation of proteins inhibiting these cellular changes (Eferl and Wagner, 2003). This expression profile is sufficient to initiate cancerous transformation of normal cells (Schutte *et al.*, 1989, Schuermann *et al.*, 1989), promote tumour expansion, and propagate late stage angiogenic and metastatic progression of many types of human cancers (Eferl and Wagner, 2003, Lopez-Bergami *et al.*, 2010).

1.1.3 Tractability of AP-1 as a viable therapeutic target

Targeting AP-1 directly may have tractability according to the following criteria, and may be a more direct and specific approach than targeting upstream activators like JNK (Takahashi *et al.*, 2013). Firstly, a causal link between AP-1 and cancers is well documented, and whilst AP-1 activity is essential for normal development (i.e. some AP-1 component knockouts are embryo-lethal) (Eferl and Wagner, 2003), AP-1 expression/activity is significantly higher in many cancer cells than in normal cells, and is different in different cancers (Lopez-Bergami *et al.*, 2010), providing opportunity for cancer-specific targeting. The lack of any mutations thus far described in AP-1 component proteins is promising for AP-1 targeting (Verde *et al.*, 2007). Though AP-1 is active in the nucleus, component proteins are expressed in the cytoplasm and actively shuttle between the cytoplasm and nucleus (Malnou *et al.*, 2007), which provides additional opportunity for their targeting. Finally, the structure of AP-1 holds promise for the design of interfering agents that will bind AP-1 specifically and with high affinity. Though different AP-1 compositions complicate the targeting of AP-1, development of specific inhibitors to each AP-1 component

protein could allow targeting of specific cancers, with therapy tailored to the specific AP-1 compositions observed in patients.

Targeting of cJun and cFos AP-1 components is particularly interesting for cancer therapy. cFos–cJun is one of the most potent transactivating compositions, and has one of the highest affinities for binding at TRE sites (Yang-Yen *et al.*, 1990, Chiu *et al.*, 1989, Ryseck and Bravo, 1991). cFos–cJun predominantly upregulates pro-proliferative protein effectors in cancers such as breast, bone, prostate and colorectal cancers (Eferl and Wagner, 2003), where cJun and/or cFos are overexpressed/hyperactivated (Lopez-Bergami *et al.*, 2010). Furthermore, cJun is particularly attractive as a target, as cFos prefers to heterodimerise (O'Shea *et al.*, 1989, Smeal *et al.*, 1989), thus requiring cJun for transcription transactivation and oncogenic transformation (Chiu *et al.*, 1988, Schuermann *et al.*, 1989), and dimerisation with cJun strongly localises cFos to the nucleus (Malnou *et al.*, 2007). Thus, agents that are capable of inhibiting the cFos–cJun PPI could be valuable for therapy of cancers featuring cFos–cJun dysregulation.

1.2 Targeting AP-1 formation: Structural determinants

1.2.1 Global Structure

AP-1 components are “leucine zipper” (LZ) proteins of the “basic-region zipper” (bZIP) family of TFs (Vinson *et al.*, 2002, Landschulz *et al.*, 1988b). They feature three main conserved domains (Figure 1.2A): the LZ domain, basic DNA-binding domain, and transcription transactivation domains. The α-helical LZ domains of two AP-1 component proteins supercoil around each other to form a parallel “coiled coil” dimeric structure (Crick, 1953, Glover and Harrison, 1995), whilst DNA-

binding domains feature basic amino acid residues that interact with DNA backbone phosphates to form a “scissor-grip” structure across the DNA major groove (Vinson *et al.*, 1989) (Figure 1.2B).

1.2.2 The Leucine Zipper Coiled Coil Domain

Because AP-1 assembly via component protein dimerisation is absolutely required for DNA binding and transactivation of transcription (Sassone-Corsi *et al.*, 1988), and the LZ CC domain drives this dimerisation (Kouzarides and Ziff, 1988), these domains may be a suitable target for AP-1 antagonists. Fortunately, numerous crystal structures and biophysical studies mean that the contributions of individual residue positions and identities to CC formation thermodynamics are reasonably well defined (Figure 1.2C), which can be exploited for the targeting of AP-1.

Crick (1953) described the supercoiling of right-handed α-helices to form left-handed CCs as occurs for the LZ domain of AP-1 (Glover and Harrison, 1995). In this model, CC domain residues of the two interacting helices interdigitate in a “knobs into holes” paradigm (Figure 1.2D). This is facilitated through distortion of the helix backbone to feature 3.5 residues per helical turn as opposed to 3.6 residues in isolated helices (Pauling *et al.*, 1951). As a result, residues in CCs display a positional and functional periodicity of seven (comprising a “heptad repeat”), and are denoted ***a*** – ***g*** on one helix, and ***a’*** – ***g’*** to distinguish those on the other helix (Landschulz *et al.*, 1988b). Each heptad of ***a***-***b***-***c***-***d***-***e***-***f***-***g*** is typically a repeat of H-P-P-H-P-P-P where H denotes hydrophobic residues and P denotes polar residues (Mason and Arndt, 2004).

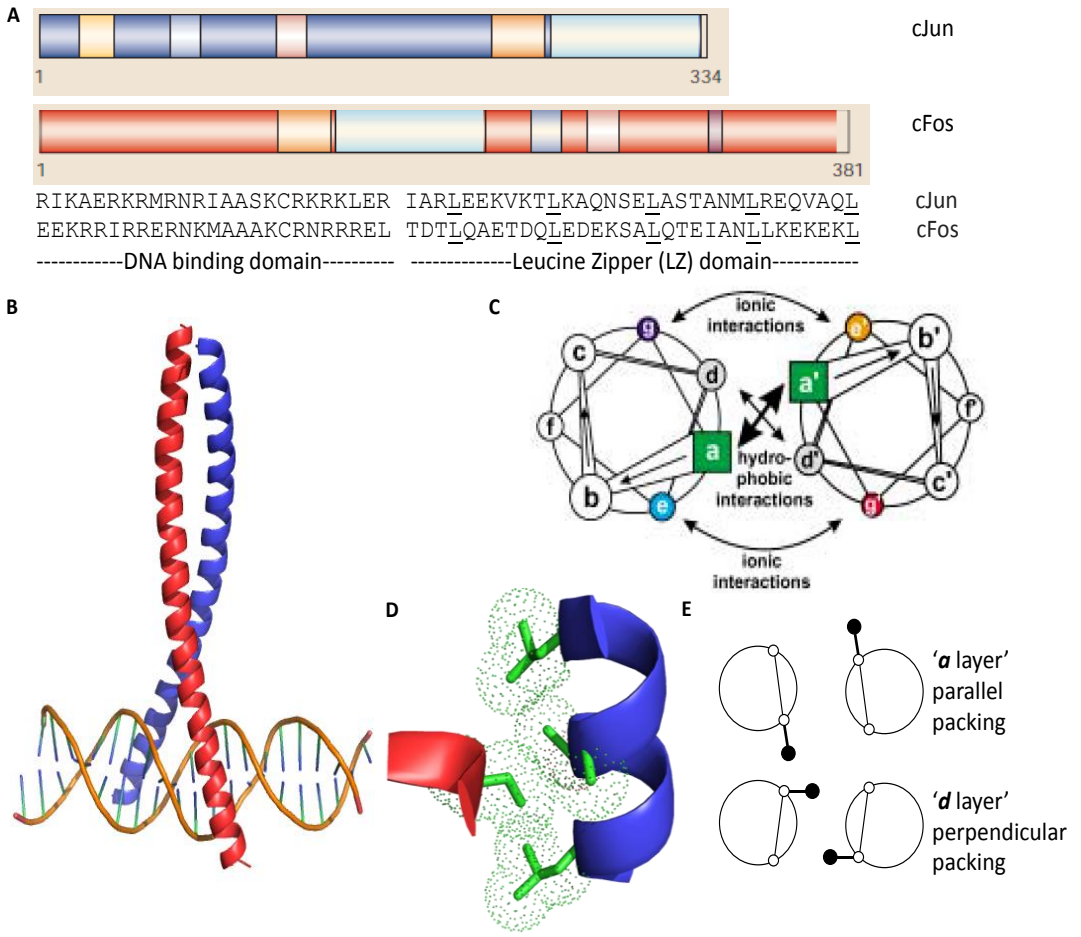


Figure 1.2: cFos–cJun AP-1 formation. A Conserved domains of cJun and cFos, adapted from source (Eferl and Wagner, 2003). cJun is coloured blue, cFos in red. Orange boxes denote DNA-binding domains, light blue boxes denote LZ domains, and mid blue and pink boxes denote transactivation domains. On cJun, the yellow box at the N-terminus denotes the JNK binding site. On cFos, the C-terminal purple box denotes the ERK binding site. Residue numbers given below diagrams. In the sequences of cJun and cFos, repeating Leu residues in the LZ domains are underlined. **B** PyMOL (Schrodinger, 2013) model of the cFos–cJun parallel CC and DNA-binding domains (PDB ID: 1FOS) assembled on the TRE binding site, from the AP-1 X-ray crystal structure (Glover and Harrison, 1995). cJun is represented as a blue ribbon, cFos as a red ribbon, the DNA backbone is in orange, and DNA bases are blue/green sticks. **C** Periodicity of residue spatial positions and functions in CCs, adapted from source (Mason and Arndt, 2004). Interhelical interactions that stabilise the coiled coil are shown. **D** PyMOL model of core "knobs-into-holes" packing (Crick, 1953) in the cFos–cJun crystal structure. An ***a*** Ile on cJun (red ribbon backbone) fits into the "hole" formed by ***d'***$_{i}$, ***g'***$_{i-7}$, ***d'***$_{i-7}$ and ***a'***$_{i}$ residues of cFos (blue backbone ribbon). **E** Packing of amino acid side chains in the core of parallel dimeric coiled coils, adapted from source (Harbury *et al.*, 1993). At ***a*** positions, side chains prefer parallel arrangements for Cα–Cβ bonds (thick black lines) relative to the peptide bond between residues (thin black lines). Conversely, at ***d*** positions, perpendicular packing is preferred.

Core ***a*** and ***d*** positions are mostly occupied by hydrophobic residues, and almost exclusively by Leucines at ***d*** positions of leucine zipper CCs such as AP-1 (Landschulz *et al.*, 1988b, Crooks, 2013). Interhelical ***a–a'*** and ***d–d'*** interactions in parallel dimers feature enthalpic van der Waals' interactions that create a solvent-excluded hydrophobic interface that winds around the inner face of the CC (Harbury *et al.*, 1993, Crick, 1953). These interactions contribute the majority of dimerisation free energy, representing the main driving force for CC formation (Acharya *et al.*, 2006). Each ***a*** position side chain packs into the cavity formed between ***d'***$_{i}$, ***g'***$_{i-7}$, ***d'***$_{i-7}$ and ***a'***$_{i}$ side chains (Figure 1.2D), and each ***d*** position packs between ***a'***$_{i}$, ***e'***$_{i}$, ***a'***$_{i+7}$ and ***d'***$_{i}$ side chains, on the partner helix (Crick, 1953). Perpendicular, parallel or acute packing side chain orientations within the core enable this interdigitation, and lead to preferences for particular amino acids (Harbury *et al.*, 1993, Harbury *et al.*, 1994). For example at ***a*** positions (the '***a*** layer') in parallel dimers, parallel packing relative to the Cα–Cα peptide bond is preferred (Figure 1.2E), and is favoured by Cβ-branched residues such as Ile and Val but not by unbranched Leu. Conversely, at ***d*** positions (the '***d*** layer'), side chains preferentially pack in perpendicular orientations relative to the Cα–Cα peptide bond (Figure 1.2E), and so Leu and other unbranched residues at the Cβ atom preferring perpendicular packing are favoured. Core–core interactions and the packing adopted contribute towards preference for CC oligomeric state adopted (Wagschal *et al.*, 1999, Harbury *et al.*, 1993), parallel or anti-parallel α-helix orientation (Monera *et al.*, 1996), and hetero- or homo-dimerisation (Zhu *et al.*, 1992). It has recently been suggested that vertical 'triads' (***d'–a–d'***) affect the stability of parallel dimers (Steinkruger *et al.*, 2012), as do 'triplet'

interactions involving two core or ***e***/***g*** positions on one helix and one of these positions on the other helix (Potapov *et al.*, 2015).

Core polar and (less frequently) charged residues, particularly at ***a*** positions, are also found at lower frequencies (Crooks, 2013). This requires unfavourable desolvation of a hydrophilic moiety and burial within a solvent-excluded space, which can be stabilised through hydrogen bonds to ***a'*** polar residues, or interaction with long polar side chains at ***g'*** positions (O'Shea *et al.*, 1991, Glover and Harrison, 1995). Polar residue preference for homotypic rather than polar-hydrophobe interactions confers preference for CC hetero- or homodimerisation (Acharya *et al.*, 2006), parallel or anti-parallel helix orientation (Oakley and Kim, 1998), and oligomeric state adopted (Lumb and Kim, 1995).

Charged residues at core-flanking ***e*** and ***g*** positions (Crooks, 2013) can participate in favourable attractive or unfavourable repulsive electrostatic interhelical $\boldsymbol{e}_{i}$–$\boldsymbol{g'}_{i-7}$ and $\boldsymbol{g}_{i}$–$\boldsymbol{e'}_{i+7}$ interactions in parallel dimeric CCs (O'Shea *et al.*, 1991), that are salt bridges (H-bonded ion pairs) or singly charged H-bonds (Smith and Scholtz, 1998). Such interactions and residues contribute to CC preference for hetero- vs. homodimerisation (Graddis *et al.*, 1993), parallel vs. anti-parallel helix orientation (Monera *et al.*, 1994), and oligomeric state adopted (Zeng *et al.*, 1997). Core-flanking residues can also form stabilising ***a***–***g'*** and ***e***–***d'*** interhelical interactions (O'Shea *et al.*, 1991, Glover and Harrison, 1995, Havranek and Harbury, 2003). Furthermore, ***e*** and ***g*** side chain carbons can form van der Waals' interactions with ***a*** and ***d*** residues, packing against the core and shielding it from solvent (O'Shea *et al.*, 1991, Harbury *et al.*, 1993).

Finally ***b***, ***c*** and ***f*** 'outerface' residues are almost completely solvent-exposed and are predominantly charged or polar (Harbury *et al.*, 1993, Crooks, 2013). Outerface residues contribute more subtly and more indirectly to CC interaction stability (O'Neil and Degrado, 1990, Mason *et al.*, 2006, Kaplan *et al.*, 2014). It is thought these residues contribute to α-helix adoption, solubility, and participation in intrahelical *i→i+4* (e.g. ***b–f*** or ***f–c***) interactions (O'Neil and Degrado, 1990, Mason *et al.*, 2006, O'Shea *et al.*, 1991). The latter may include salt bridges and singly charged H-bonds (Smith and Scholtz, 1998), cation–π interactions between charged and aromatic residues (Andrew *et al.*, 2002, Slutsky and Marsh, 2004), and polar-apolar residue van der Waals' interactions (Andrew *et al.*, 2001). Less frequent ***b–e*** and ***c–g*** interactions, perhaps indirectly affecting interhelical interactions involving core-flanking residues, have also been observed (Kohn *et al.*, 1997, O'Shea *et al.*, 1991).

1.3 Targeting AP-1: the advantages of peptides as antagonists

AP-1 structural knowledge enables the design of antagonists, which may be either "small molecule" drugs (generally ≤500 Da in size), or biologics (peptides or proteins, generally ≥500 Da). For disease therapy, small molecules have been the conventional choice for a number of reasons (Craik *et al.*, 2013). For example, small molecule synthesis is cheaper than biologic synthesis or recombinant expression. Small molecules can be highly bioavailable across biological membranes, and stable in extracellular fluids. Conversely, biologics are often too large to traverse cell membranes easily, and are unstable in *in vivo* environments due to protease susceptibility, poor solubility and rapid systemic clearance, preventing desirable oral

administration. Small molecules are non-immunogenic, whilst biologics can contain epitopes recognised by T and B cells (Fosgerau and Hoffmann, 2015).

Despite this, biologics exhibit beneficial attributes that make them more attractive for modulation of disease-involved PPIs than small molecules, and further are amenable to a wide variety of modifications to overcome limitations in their therapeutic viability (Liskamp *et al.*, 2011). Biologics have a major advantage over small molecules in that their larger size allows formation of more points of interaction with the target, to increase specificity over "off-targets", interaction with which may generate toxicity. PPIs like that in AP-1 which feature large, hydrophobic surfaces, are non-ideal binding sites for small molecules: small molecules are too small to interfere with such PPIs significantly, and form too few specific (i.e. non-hydrophobic) interactions for selective targeting due to their small size (Lo Conte *et al.*, 1999, Corbi-Verge and Kim, 2016). Furthermore, small molecule development can be difficult without experimentally-confirmed structural information on binding modes of previously-derived, structurally-related small molecules. Indeed, there are currently no crystal or NMR structures of small molecule–AP-1 complexes in the Protein Data Bank (PDB). Conversely, previously-derived, simplistic relationships between α-helix residue properties/positions and coiled coil stability can be exploited as design rules for α-helical peptide antagonists that sequester AP-1 component proteins into non-functional complexes via interactions with their LZ CC domains (Mason *et al.*, 2006). The major advantage of this approach is the specificity for AP-1 antagonism that can be achieved through design of complementary ***a***–***a'***, ***d***–***d'***, and ***e***–***g'***/***g***–***e'*** coiled coil interactions. Moreover, antagonist characterisation enriches our understanding of CCs such that future antagonists with similar binding

modes can be better rationally designed. Finally, intrinsic features of α-helical peptides make them attractive for development into therapeutic agents: typically they do not fit well into protease active sites (Madala *et al.*, 2010), and stabilised helices may have some ability to cross biological membranes (Sun *et al.*, 2013).

1.3.1 Previous AP-1 small molecule antagonists and their shortcomings

In agreement with the view that PPIs represent difficult targets for small molecules, currently there are few small molecules targeting AP-1 that are under development (Yap *et al.*, 2012, Ye *et al.*, 2014). These include natural products such as resveratrol and curcumin and their derivatives, and synthetic designed molecules. However, issues with target specificity stem from the planar, aromatic nature of these molecules. For example, resveratrol interferes with MAPK signalling upstream of AP-1 (Yu *et al.*, 2001), whilst direct AP-1 inhibitors curcumin and momordin I, which are reported to block DNA binding, show significant cross-reactivity with heterologous proteins. Synthetic small molecule T-5224 derived from a cyclic peptide, on the other hand, displays some selectivity over other transcription factors (Aikawa *et al.*, 2008). Nevertheless, T-5224 targets the DNA-binding domain of AP-1, which shares very high sequence homology across other bZIP TFs (Fujii *et al.*, 2000), such that targeting specificity may be an issue. Further, specificity for particular compositions of AP-1, essential given that some compositions are anti-oncogenic in certain situations (Eferl and Wagner, 2003), is currently lacking.

1.3.2 Previous AP-1 peptide antagonists and their shortcomings

Previously, a number of AP-1 antagonist biologics have been described. These have generally been competitive inhibitors that sequester AP-1 component proteins

into complexes either incapable of transactivating transcription of tumourigenesis-related genes ("quenching"), or unable to bind AP-1 recognition sites in AP-1 target gene promoters ("squelching"), thus preventing functional AP-1 assembly.

Initially, truncation of native AP-1 component proteins generated relatively lengthy antagonist proteins, which nevertheless demonstrated the potential of biologics for AP-1 inhibition and the tractability of AP-1 as a therapeutic target. These proteins include "Δ9", a 149 residue protein created by cJun N-terminal transactivation domain truncation, which reduced AP-1-driven gene transcription to levels similar to basal AP-1 activity and inhibited cancerous transformation (Lloyd *et al.*, 1991). TAM67 is an alternative 210mer protein created by similar cJun N-terminal truncation, which inhibited transformation by cJun or cFos (Brown *et al.*, 1994) and further reduced xenograft tumour growth, anchorage-independence, and invasiveness when transfected into diverse cancer cell lines (Jin *et al.*, 2007, Maritz *et al.*, 2011, Leaner *et al.*, 2009). Finally, various cFos mutant proteins lacking the LZ domain or DNA-binding domain have demonstrated prevention of AP-1-mediated transformation (Wick *et al.*, 1992). However, targeting specificity is an issue, as these proteins, like their native counterparts, interact with multiple AP-1 component proteins (Thompson *et al.*, 2002).

More recently, antagonist development has focused on shorter α-helical peptides generated from the LZ CC domains of AP-1 component proteins. In addition to the lending of greater structural information of these domains to antagonist design as discussed above, LZ CC domains generally exhibit lower homology across diverse bZIP TFs than, for example, their DNA-binding domains (Vinson *et al.*, 2002,

Fujii *et al.*, 2000), aiding the design of more specific antagonists. Shorter peptides include the "A298V superzipper", a 46mer peptide based on an Ala to Val mutant of the cJun LZ domain, which displayed preferential dimerisation with cJun *in vitro* rather than with itself, though also dimerised with cFos with similar affinity (Bains *et al.*, 1997). More recently, a peptide-DNA conjugate ("C2ds") has been described that combines a 36mer cFos LZ CC domain with an 18 bp dsDNA molecule containing the sequence TCAT recognised by cJun (Pazos *et al.*, 2015). This molecule demonstrated submicromolar AP-1 inhibition by electrophoretic mobility shift assay and fluorescence anisotropy *in vitro*. Though displaying attractive affinities, as for longer protein antagonists the use of predominantly native CC domains can undesirably lead to non-specificity for particular AP-1 compositions. More recently, the 42mer "Jun-d1" peptide was designed against the cJun LZ CC domain using an *in silico* interaction scoring algorithm derived from peptide interaction microarray data, and was shown to bind cJun *in vitro* with attractive nanomolar affinity and specificity over some other bZIP TFs (Potapov *et al.*, 2015). Finally, 32 – 37mer antagonist peptides with desirable specificity for cJun over cFos have been successfully generated from extensive mutation of the cFos LZ domain (Crooks *et al.*, 2011, Worrall and Mason, 2011).

Though some of the antagonists described above display attractive inhibition of AP-1 formation and/or activity, they are generally suboptimal for therapeutic administration. This is in part due to their relatively large size compared to small molecules, particularly for the early protein antagonists, and in part to their intrinsic physicochemical properties as unmodified peptides, both of which are likely to severely limit bioavailability, and make antagonists vulnerable to protease digestion and immune system reactivity (Fosgerau and Hoffmann, 2015). Antagonist

downsizing and/or modification may, however, circumvent these issues to create viable therapeutic molecules (Craik *et al.*, 2013, Liskamp *et al.*, 2011).

There is thus a need to identify novel specific, short peptide AP-1 antagonists with more attractive attributes for future human cancer therapy. Identification of novel peptides can be achieved by individual peptide characterisation as generally has been performed previously, or using peptide library display and selection systems for much higher throughput.

1.4 Identifying and developing peptides towards therapeutic agents

1.4.1 Peptide library display and selection systems for peptide identification

Peptide library display and selection systems are high throughput, rapid techniques for the isolation of peptides binding to a target (Baxter *et al.*, 2014). They feature generation of a nucleic acid library of peptide variants, expression such that there is a direct "phenotype-genotype linkage" between peptide ("phenotype") and encoding material ("genotype"), and then screening for peptides able to bind the target. Selective pressure can be applied to generate competition between target-binding peptides to select only the high affinity binders of the target. Linked genotype can then be sequenced to provide identification of "hits" for further development. In addition to screening for affinity, other selective pressures can be applied to select for peptides with other desirable properties. Selected peptides can then be assayed for modulation of the target's activity.

Library construction can be achieved by a number of approaches. Traditional approaches are mainly PCR-based, using oligonucleotide primers carrying

randomised codons. “Megaprimer” and “overlap extension” methods are commonly employed (Ke and Madison, 1997, Ho *et al.*, 1989). In the megaprimer method, the wild type gene is mutated with a mutagenic primer to generate a long “megaprimer”, which subsequently is used to amplify the full length mutant gene. In overlap extension, hybridisation of a long mutagenic primer with a short primer sharing a complementary region is followed by extension of each strand using the other as a template. Alternatively, oligonucleotide libraries can be synthesized directly through coupling of mono-, di- or trinucleotide phosphoramidites, the latter being particularly attractive as these represent DNA codons (Virnekas *et al.*, 1994). Unfortunately, these techniques can be complex, inefficient and expensive for highly diverse libraries (Ashraf *et al.*, 2013).

Newer construction techniques include ProxiMAX randomisation (Ashraf *et al.*, 2013) (Figure 1.3) and “Slonomics™” (Van den Brulle *et al.*, 2008). ProxiMAX features blunt-end ligation of “donor” (bearing library codons) and “acceptor” dsDNA oligonucleotides, PCR amplification of the ligation product, and then double-strand blunt-end digestion to remove the donor and leave the library codon on the growing acceptor, ready for the next codon addition. Slonomics is an analogous method that uses compatible single strand overhangs instead of blunt-ends for ligation. The benefit of these techniques over traditional PCR approaches is that multiple contiguous sites can be fully randomised with far fewer primers, the number of which otherwise dramatically increases with randomisation extent and site number and can be inhibitory for highly diverse libraries. ProxiMAX is particularly attractive as it requires fewer primers for randomisation than Slonomics (Ashraf *et al.*, 2013).

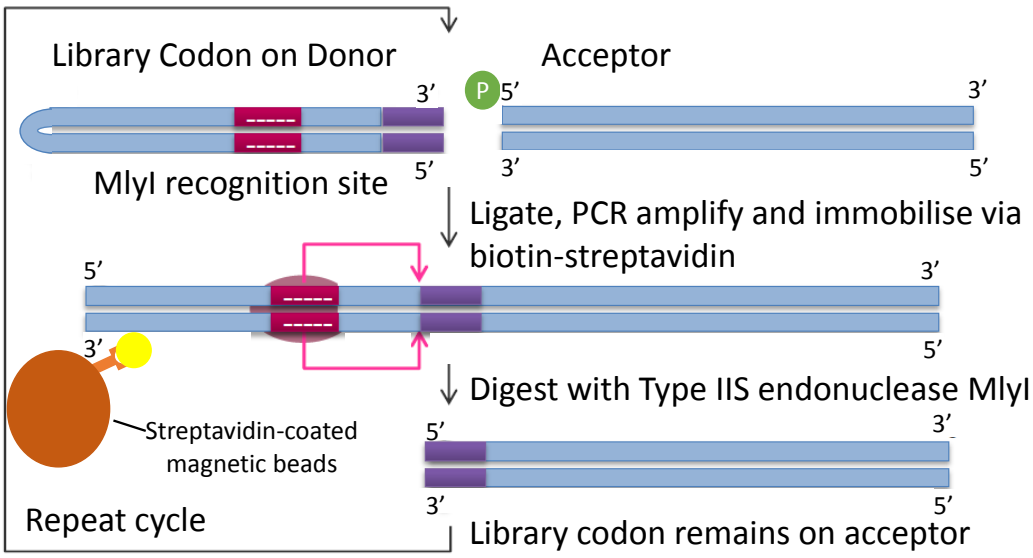


Figure 1.3: Library construction by ProxiMAX randomisation. A library codon (purple) encoding an amino acid is delivered to a growing library "acceptor" on a dsDNA hairpin oligonucleotide "donor". Directional 'blunt-end' ligation of the donor and acceptor dsDNA molecules is catalysed by T4 DNA ligase making use of the acceptor's 5' phosphate (green circle), and the ligation product is selectively amplified by PCR introducing a biotin tag. The amplicon is immobilised via streptavidin-biotin interaction. Digestion by Type IIS restriction endonuclease MlyI (pink oval), acting downstream of its recognition site (dark pink box) in the donor, removes the donor from the amplicon to leave the library codon on the acceptor. Repeated cycles of codon addition sequentially add DNA codons to the growing library. Figure adapted from source (Ashraf *et al.*, 2013).

To screen libraries against a target, display and selection systems utilising cell-free *in vitro* library expression ("*in vitro*" systems) (Figure 1.4) feature either DNA library encoding, such as CIS-display (Odegrip *et al.*, 2004), covalent DNA display, and *in vitro* compartmentalisation displays; or RNA encoding, such as in ribosome display, and mRNA display (Baxter *et al.*, 2014). Typically, libraries are expressed using ribosomal machinery from cell lysate or purified components (Shimizu *et al.*, 2005), and then linked to expressed peptide. This is achieved via display of library peptides on the surface of: DNA-binding proteins (CIS display/covalent display), stalled

ribosomes (ribosome display), microbeads via antibody–epitope binding within water droplets in an oil emulsion ("*in vitro* compartmentalisation"), or on the ribosome inhibitor puromycin (mRNA display). Peptides are incubated with an immobilised target *in vitro*, low affinity binders are removed by washing, and competition for the target selects the highest affinity binders.

Display and selection systems featuring library expression in cells (and thus classed as "*in vivo*" or "*in cellulo*" systems) (Figure 1.4) generally make use of DNA library encoding, such as Protein-fragment Complementation Assays (PCAs) (Pelletier *et al.*, 1999), or phage, yeast two-hybrid, and bacterial or yeast/mammalian cell surface displays (Baxter *et al.*, 2014). Libraries are generated *in vitro*, transformed/transfected into individual host cells, and expressed by ribosomal machinery. Library peptides are displayed on the surface of cell surface proteins (bacterial/yeast/mammalian display) or bacteriophage coat proteins (phage display) and are incubated with a target *in vitro* as described above for systems using cell-free expression. Alternatively, library peptides are co-expressed in cells with the target and library–target interaction generates a measurable readout. In PCAs, refolding of split essential proteins fused to library and target, such as dihydrofolate reductase (DHFR-PCA), gives a detectable readout, which in the case of DHFR-PCA is nucleotide synthesis and cell growth (Pelletier *et al.*, 1999), whilst refolding of split fluorescent proteins allows fluorescence sorting of cells (Kerppola, 2006). Finally, in yeast two-hybrid systems, library–target interaction allows Gal4 transcriptional activator-driven transcription from reporter genes.

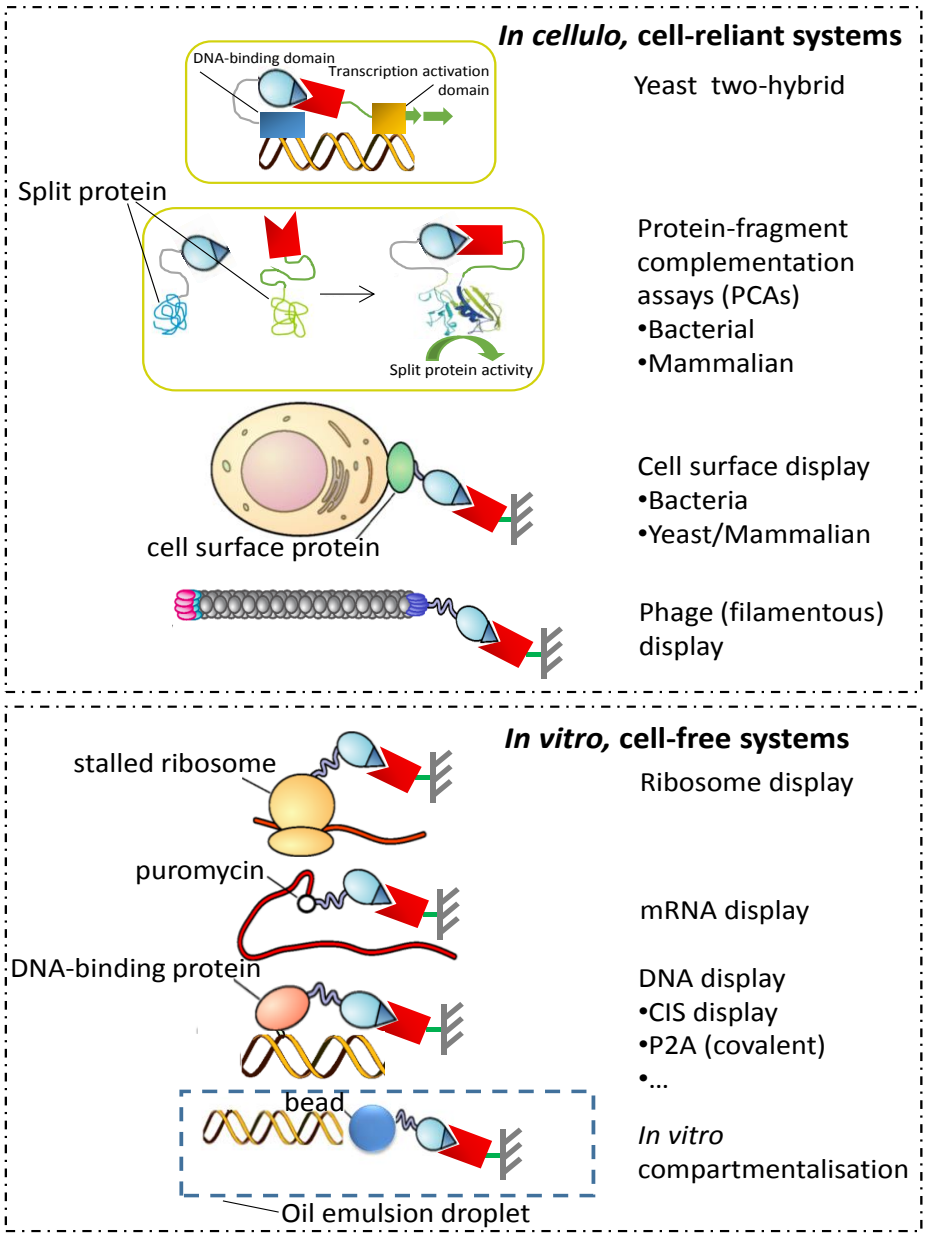


Figure 1.4: Library display and selection systems. Schematic of library display for selection against a target in commonly used library display systems which make use of cellular expression of libraries (top panel), or *in vitro* cell-free expression (bottom panel). The target of interest against which libraries are screened is shown as a red rectangle with triangular cut-out. For the majority of display systems shown, the target is immobilised on solid support (grey dashed lines) via a linker (green) and libraries are panned against it *in vitro*. Library-encoding DNA is shown as a yellow ribbon, or mRNA as a red ribbon, displayed peptide is shown as a light blue oval, with its target-binding site shown as a dark blue triangle. Linkers between displayed peptide and the molecule on which it is displayed are shown as grey ribbons. Cell walls are shown as yellow rectangles. See text for a more detailed explanation of display systems. Figure adapted from source (Sergeeva *et al.*, 2006).

In vitro and *in vivo* systems have their strengths and weaknesses, some of which overlap. *In vitro* systems allow the expression of peptide libraries of much higher diversities ($\leq 10^{14}$ peptides) than *in vivo* systems ($\leq 10^{11}$ peptides) due to the limitations to transformation/transfection efficiencies of the latter (Baxter *et al.*, 2014). This enables greater sampling of peptide sequence space, to increase the chance of identifying peptide "hits". False positives are a rarer occurrence in *in vivo* systems where the stringency of selection is higher (Remy and Michnick, 1999), whereas *in vitro* systems can suffer from inadequate removal of non-specific off-target binders (Nieuwlandt, 2000). Sensitivity of *in vivo* systems to weaker target binders (which may nevertheless generate useful antagonists with further affinity maturation) can be much better than *in vitro* systems (Remy and Michnick, 1999). Selection *in vivo* has the advantage of simultaneously selecting for solubility in the cell environment, non-aggregation, stability within a reducing environment, protease resistance, and ability to outcompete endogenous competitors, both for the target, and the library (Pelletier *et al.*, 1999, Mason *et al.*, 2007b). Finally, *in vitro* systems benefit from increased control and easier manipulation of additional selection stringencies such as incubation temperature, washing harshness, altering pH or denaturant concentration, adding competitors (such as heterologous targets or previously selected target-binders), and adding proteases (Eldridge *et al.*, 2009).

An example of an attractive *in vitro* system is CIS display, which takes advantage of a bacterial plasmid replication initiation protein RepA to provide a phenotype-genotype linkage for affinity selection (Odegrip *et al.*, 2004). RepA displays *cis*-activity, the high-fidelity binding of RepA to an *oriR* (origin of replication) on the same DNA molecule as RepA is expressed from (Masai *et al.*, 1983), such that

library peptides expressed in fusion with RepA are linked to their encoding DNA. CIS display combines screening of hugely diverse libraries ($\leq 10^{14}$ peptides) with ease of manipulation of selection pressures, and the greater stability of DNA for library encoding relative to RNA used in mRNA and ribosome displays.

The bacterial DHFR-PCA of Pelletier *et al.* (1999) is a particularly attractive *in cellulo* system. Library peptides are genetically fused to a rationally-designed fragment of murine DHFR (mDHFR1) whilst the target is fused to the complementary fragment, mDHFR2. *E. coli* harbouring library-mDHFR1 and target-mDHFR2 DNA are grown in minimally nutritious medium, and genomic DHFR production of crucial nucleotide precursor tetrahydrofolate is suppressed by the antibiotic trimethoprim. Library and target molecules are specifically expressed from *lac* promoters in host cells through alleviation of transcriptional repression by the *lac* repressor using isopropyl-β-D-1-thiogalactopyranoside (IPTG). If library peptides interact with the target, the mDHFR fragments can refold and form active mDHFR, to allow nucleotide synthesis and cell growth. Thus, growing bacteria contain target-interacting library peptides whose encoding DNA can be sequenced to identify them. DHFR-PCA features selection of desirable therapeutic-like properties of peptides as well as affinity in a rapid and facile manner due to the ease of bacterial cell growth.

1.4.2 Developing peptides into therapeutic agents

A small number of unmodified peptides have been successfully administered therapeutically, with enfuvirtide, a 36mer peptide that blocks HIV-1 fusion to CD4+ T cells, being an example that is often quoted. However, its protease susceptibility in systemic fluids and rapid renal clearance necessitates large doses (Kilby *et al.*, 2002),

thus demonstrating the difficulty in using unmodified peptides as therapeutic agents. Increasingly, peptides are progressing through clinical trials in modified forms (Fosgerau and Hoffmann, 2015), aided by the amenability of peptides to modification and the diverse range of modifications available (Liskamp *et al.*, 2011), which generate molecules that "bridge the gap" between small molecule bioavailabilities/stabilities and biologic specificities (Craik *et al.*, 2013).

Peptides can be converted into more drug-like "peptidomimetics" using non-natural/alternative amino acid, structural constraint/cyclisation, conjugation and truncation strategies (Pelay-Gimeno *et al.*, 2015). These are generally applied to peptides after library selection, though non-natural and D-amino acids (Kawakami and Murakami, 2012, Schumacher *et al.*, 1996), helix constraints (Heinis *et al.*, 2009), conjugation and truncation (Crooks *et al.*, 2011) can be applied pre-selection, and the range of pre-selection modifications that can be made is growing rapidly. Furthermore, combining modifications can provide synergistic improvement of peptide therapeutic properties.

Natural α-amino acids can be substituted with those bearing side chains that are highly derivatized, and/or featuring non-natural chemistries, to overcome limitations to the strength and complementarity of α-amino acid interactions (Young and Schultz, 2010). Alternatively, for the mimicry of α-helical peptides, partial substitution for amino acid mimetics like β-amino acids can generate functional, conformationally similar helical peptides that are more resistant to protease digestion (Johnson and Gellman, 2013). D-α-amino acid substitution, and the related "retro-inverso" strategy, replace natural L-enantiomer α-amino acids with D-

enantiomers, and reverse the D-α-amino acid sequence from N- to C-terminus to arrive at a 'mirror image' of the L-enantiomer respectively (Goodman and Chorev, 1979). This decreases susceptibility to proteases and immune responses evolved to recognise L-α-amino acids (Yamada and Kera, 1998, Benkirane *et al.*, 1993). Alternatively, peptide backbones can be replaced with moieties and chemistries more resistant to proteases. For α-helices, functional groups capable of adopting the necessary Φ and Ψ angles for effective structural mimicry include benzene ring derivatives featuring amides, ureas, and azoles (Lanning and Fletcher, 2013).

Helix constraints are non-natural modifications to α-helical peptides that cyclise two amino acids and thus the peptide backbone, to promote helicity through reduction of the entropic cost to α-helix adoption (Rao *et al.*, 2013). Constraints can also provide significant protection from protease digestion and may improve cell penetrance (Schafmeister *et al.*, 2000, Sun *et al.*, 2013). Constraint modalities include lactam constraints (Harrison *et al.*, 2010), aryl halides (Timmerman *et al.*, 2005), hydrocarbon chains (Schafmeister *et al.*, 2000), and metal chelation to force helicity (Kelso *et al.*, 2004), and can be inserted at suitably spaced positions encouraging helical conformations, with *i→i+3*, *i→i+4*, *i→i+7* and *i→i+14* being close/in line with each other. A related strategy is that of hydrogen bond surrogates, where covalent hydrocarbon chains replace terminal backbone hydrogen bonds that stabilise a helical peptide. A benefit over helix constraints is that there may be less interference with residue side chains involved in target interaction (Chapman *et al.*, 2004). Finally, termini macro-cyclisation is also an attractive strategy (Craik *et al.*, 2013).

Peptides can be conjugated to protein transduction domains to enhance

cellular uptake. These include cell-penetrating peptides (CPPs) such as the HIV-1 Transactivator of transcription (Tat) peptide, which are capable of crossing biological membranes by various mechanisms (Debaisieux *et al.*, 2012, Herce *et al.*, 2014). Conjugation to carrier molecules such as polyethyleneglycol (PEG) slows renal clearance to improve tissue uptake, improves proteolytic stability, and lowers immunogenicity, though antibodies to PEG and deposition in tissues may be more of an issue than previously realised (Qi and Chilkoti, 2015). Finally, truncation may also improve cell penetration potential (Crooks *et al.*, 2011, Agrawal *et al.*, 2016).

1.5 Thesis Aims

The AP-1 transcription factor represents an attractive and as yet 'undrugged' therapeutic target for a variety of cancers featuring its dysregulation, warranting further exploration of novel antagonists of its activity. There is considerable potential for α-helical peptides to be high affinity and high specificity antagonists rather than traditional small molecule drugs, but previously derived peptides are suboptimal for therapeutic application due to their instability, non-specificity, and/or large size.

As such, this thesis had two major aims. The first was to explore helix constraint modifications of high affinity, yet small, AP-1 antagonist peptide "FosW", which targets cJun to prevent cFos–cJun formation (Mason *et al.*, 2006). It was hypothesized that helix constraint would entropically pre-organise FosW variants for improved binding to cJun, which should allow truncation of these constrained peptides down to a smaller and more therapeutically attractive size, whilst retaining sufficient binding affinity despite loss of interaction points with the target. Such an approach was recently shown to be effective for a cFos antagonist peptide (Rao *et*

al., 2013). Constrained and truncated FosW derivatives were analysed for retention of coiled coil formation with cJun and binding free energy using *in vitro* biophysical characterisation techniques.

The second aim of this thesis was to identify peptide antagonists of cFos–cJun AP-1 using a novel library screening technique involving combination of CIS display (Odegrip *et al.*, 2004) and bacterial DHFR-PCA (Pelletier *et al.*, 1999). It was hypothesized that this combination (hereafter referred to as "CIS→PCA") could exploit potential synergy between these systems for selection of peptides from highly diverse libraries (not easily achievable with PCA) with *in cellulo* refinement of desirable, drug-like properties (not easily achievable using CIS display). Libraries of peptides shorter than FosW were designed and screened for cJun binders, with the hypothesis that shorter peptides could retain sufficient binding affinity for cJun to encourage further development towards a therapeutic agent in the future. Highly diverse peptide libraries were constructed by the novel ProxiMAX randomisation technique (Ashraf *et al.*, 2013) and screened by CIS display. Peptide hits were then transferred to DHFR-PCA, selected for cJun binding *in cellulo*, and resultant peptides synthesized and characterised using *in vitro* biophysical characterisation techniques.

The following chapters describe Materials and Methods (Chapter 2), helix constraint and truncation of FosW (Chapter 3), CIS→PCA selection of novel AP-1 antagonists (Chapters 4 and 5), and a general discussion and suggestions for future work (Chapter 6).

Chapter 2 – Materials and Methods

2.1 Materials

All materials were purchased from Sigma-Aldrich (Gillingham, UK) unless stated otherwise. Solutions are aqueous unless specified otherwise. Reaction setups are given for 1X reactions, which were scaled up as necessary.

2.2 Library Designs

Three libraries ("2.5h", "3.5h" and "3.4hFosW") were constructed based on truncation and re-randomisation of FosW (Mason *et al.*, 2006). Two additional libraries ("12mer" and "16mer") featured full randomisation of 12 or 16 amino acid positions. Library designs, including randomisation choices (incorporated in equal proportions, except aromatic residues introduced at 1.67 % each), lengths and total sequence diversities, are displayed in Figure 2.1.

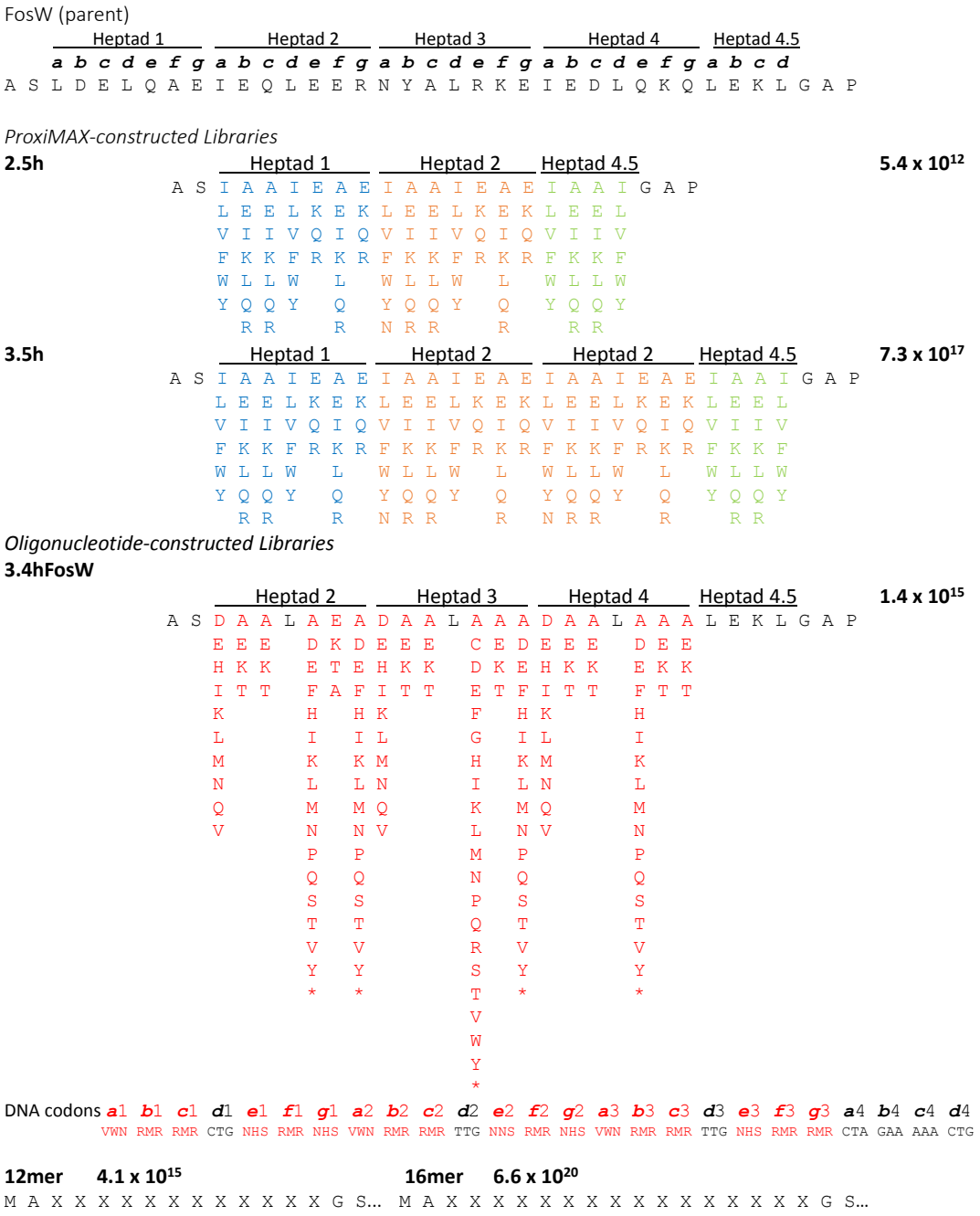


Figure 2.1: Peptide library designs. 2.5h and 3.5h libraries were built in a modular fashion from Heptad 1, 2 and 4.5 cassettes constructed using the ProxiMAX codon-ligation approach (Ashraf *et al.*, 2013). Within each cassette, amino acids were incorporated at equal frequencies, except aromatic residues F, W and Y, which were introduced at a total frequency of 5 % (split equally). Heptad positions are in bold italics. The 3.4hFosW, 12mer and 16mer libraries were built by PCR using degenerate oligonucleotides. X denotes any of the 20 natural amino acids and one stop codon (asterisk), encoded by NNN codons (N = A/C/G/T). DNA degenerate bases are: V = A/C/G, W = A/T, N = A/C/G/T, R = A/G, M = A/C, H = A/C/T, S = G/C. Library theoretical diversities are stated in bold text.

2.3 ProxiMAX-CIS→PCA overview

Chapters 4 and 5 describe CIS→PCA for the isolation of peptides with affinity for cJun (Figure 2.2). The 2.5h and 3.5h library variable regions were constructed in "cassettes" ("Heptad 1", "Heptad 2" and "Heptad 4.5") by the ProxiMAX method (Ashraf *et al.*, 2013), which were then ligated in a step-wise fashion (Figure 2.3), whilst 3.4hFosW, 12mer and 16mer libraries were constructed by standard PCR. To allow CIS display screening, RepA and associated coding sequences (Appendix Figure 2.1) were attached to 2.5h and 3.5h library variable regions. This was achieved by sticky-ended ligation following AscI endonuclease digestion at the 3' end of libraries, and MluI digestion at the 5' end of the RepA coding sequence. The resultant ligation product was then PCR amplified to add the *tac* promoter. CIS display was performed as described previously (Odegrip *et al.*, 2004) with minor proprietary modifications. DNA recovered from the 4th round of selection of 2.5h and 3.5h libraries was PCR amplified, and the C-terminal AscI site re-activated in a subsequent PCR, to allow cloning into the PCA library vector via NheI/AscI sites. Deep sequencing of round 4 DNA allowed estimation of cloning quantities for sufficient sequence over-representation to cover the selected peptide diversity in the following PCA. PCA was performed as previously described (Pelletier *et al.*, 1999) with minor modifications. Finally, peptides interacting with cJun within bacterial cells were identified by Sanger sequencing, and were synthesized, purified, and characterised.

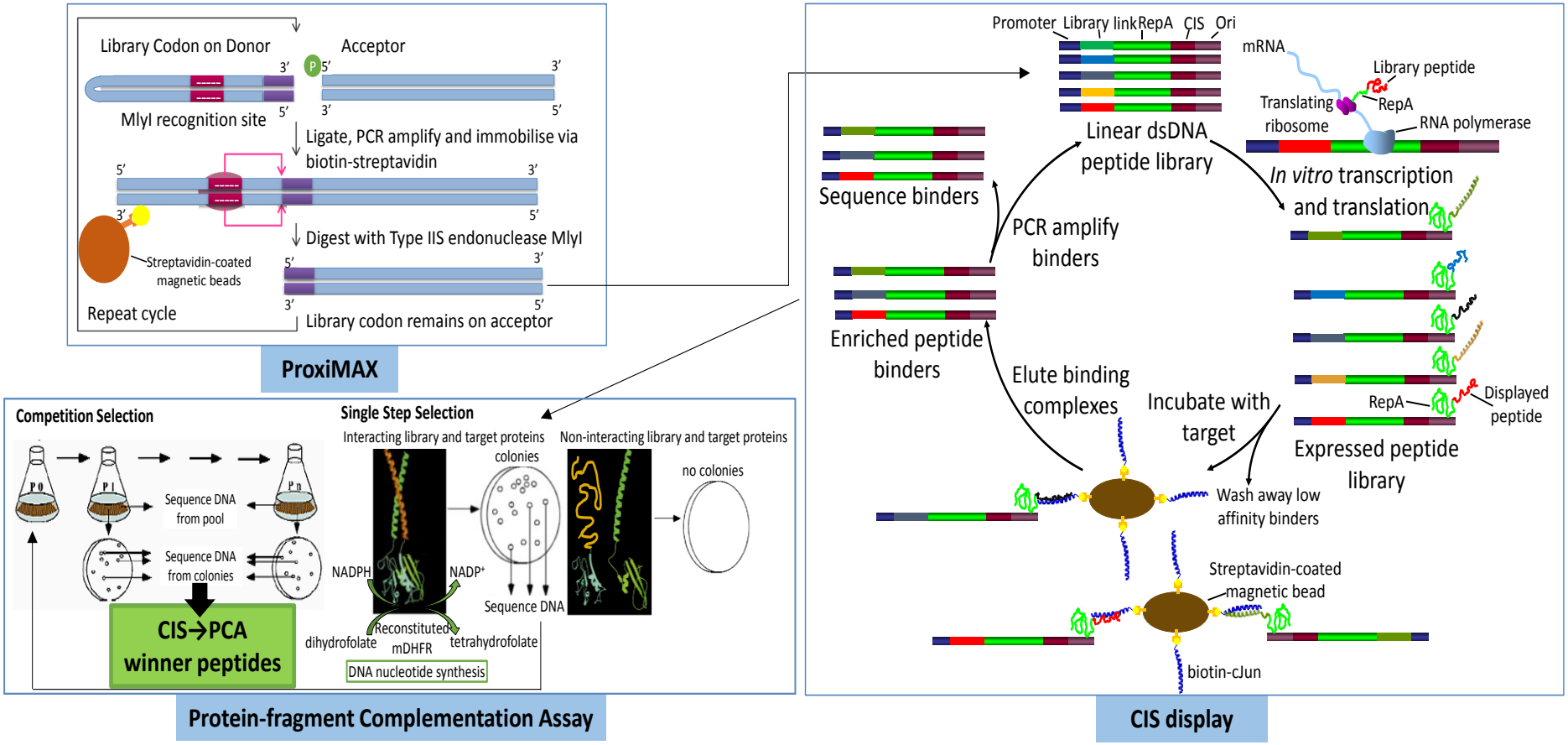


Figure 2.2: ProxiMAX-CIS→PCA. Library construction by ProxiMAX, and selection of cJun binding peptides by sequential CIS display and Protein-fragment Complementation Assay (PCA). Black arrows indicate the path of progression through ProxiMAX-CIS→PCA. ProxiMAX schematic adapted from source (Ashraf *et al.*, 2013), CIS display schematic adapted from source (Odegrip *et al.*, 2004), and PCA schematic adapted from source (Pelletier *et al.*, 1999). For further details of ProxiMAX, CIS display and PCA, see Figure 1.3 and section 1.4.1.

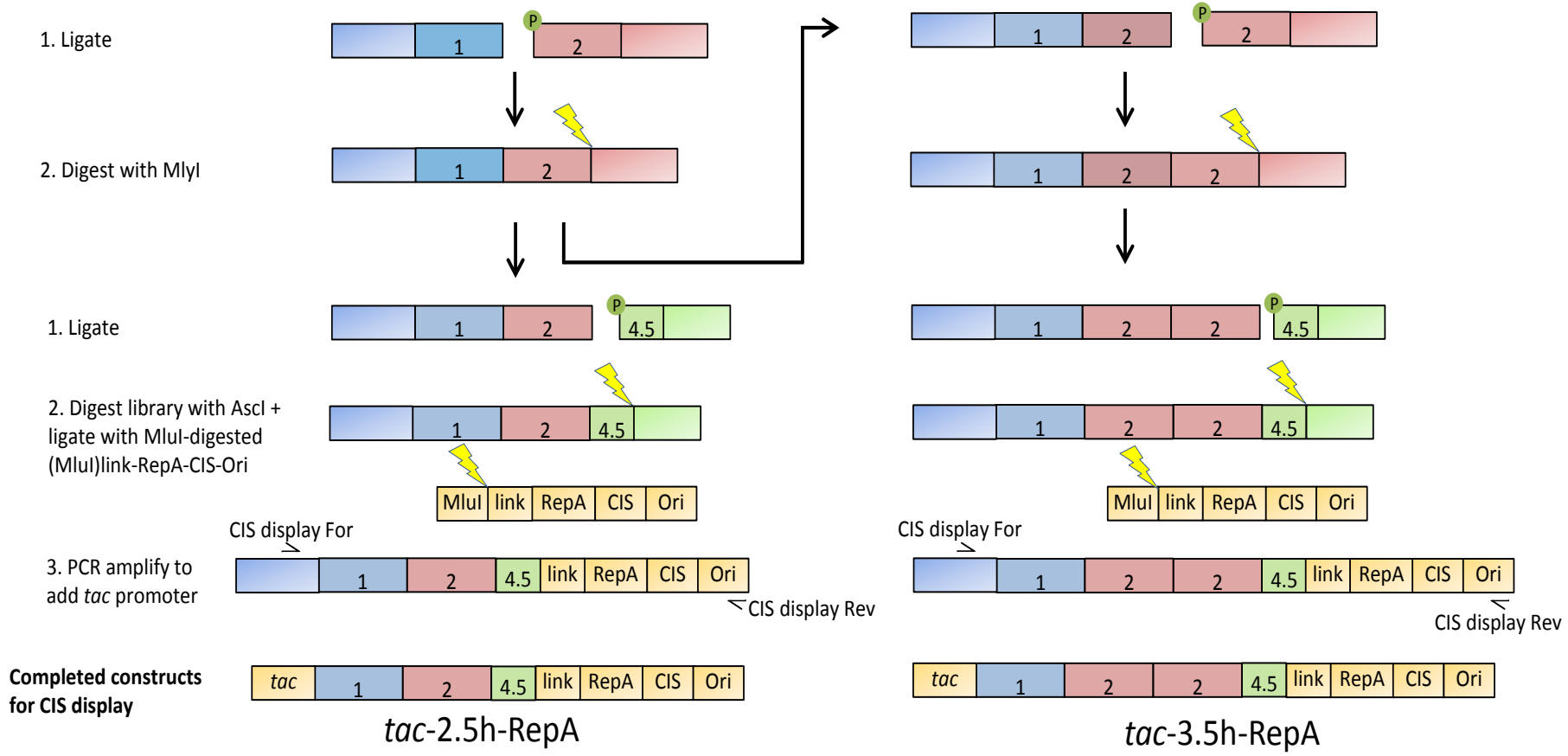


Figure 2.3: 2.5h and 3.5h library builds from Heptad cassettes. Schematic for Heptad cassette ligation to form complete library variable regions 2.5h and 3.5h, and attachment of CIS display sequences to form constructs ready for CIS display selection. Green circles indicate 5' phosphate groups, yellow lightning bolt icons represent dsDNA endonuclease restriction, and half arrows indicate PCR primers (names indicated).

2.4 Molecular Biology

2.4.1 Materials

Solutions used for Molecular Biology and CIS display are displayed in Table 2.1. All ddH_2O was PCR-grade sterilised water.

2.4.2 Donor hairpin preparation

Cycle C1, C2 and C3 donor hairpins (LGC Biosearch Technologies, CA, USA) for library codon additions during Heptad cassette construction were self-annealed by heating to 95 °C in EB buffer (Table 2.1) in a water bath and allowing to cool to 25 °C.

2.4.3 Ligations

Ligations were typically performed as outlined in Table 2.2, using T4 DNA ligase (Thermo Scientific, Loughborough, UK) for 15 min at 25 °C, and analysed by agarose gel electrophoresis for formation of expected species (see 2.4.6).

For ligation of library codons to growing Heptad cassettes, DNA 1 in Table 2.2 was a C1, C2 or C3 donor hairpin mixture (0.12 μM final), and DNA 2 was a C1, C2, or C3 Acceptor (0.02 μM final) (acceptor sequences in Appendix Figure 2.2).

Table 2.1: Molecular Biology/CIS display solution compositions. TE was bought from Calbiochem (Nottingham, UK), and TAE and TBE from Invitrogen (Loughborough, UK).

Buffer/Solution	Components
EB (QIAGEN)	10 mM Tris-Cl pH 8.5
TE	10 mM Tris, 1 mM EDTA, pH 8.0
TAE	40 mM Tris-acetate, 1 mM EDTA, pH 8.3
TBE	100 mM Tris base, 1 mM EDTA, 90 mM boric acid, pH 8.4
2X B&W	10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl
QIAGEN GelPilot or Bioline DNA loading buffer	(proprietary): contains bromophenol blue/xylene cyanol/orange G or bromophenol blue, glycerol
Fermentas GeneJET Resuspension Solution	(proprietary), 0.1 mg/ml RNaseA
Fermentas GeneJET Lysis Solution	(proprietary): contains SDS, NaOH
Fermentas GeneJET Neutralisation Solution	(proprietary): contains GdnCl
Fermentas GeneJET Wash Solution	(proprietary): contains 64 % ethanol
QIAGEN Endofree Maxi P1 buffer	50 mM Tris-HCl, 10 mM EDTA, 0.1 mg/ml RNaseA, pH 8.0
QIAGEN Endofree Maxi P2 buffer	200 mM NaOH, 1 % SDS
QIAGEN Endofree Maxi P3 buffer	3 M KAc, pH 5.5
QIAGEN Endofree Maxi ER buffer	(proprietary)
QIAGEN Endofree Maxi QBT buffer	750 mM NaCl, 50 mM MOPS, 15 % isopropanol, 0.15 % Triton X-100, pH 7.0
QIAGEN Endofree Maxi QC buffer	1 M NaCl, 50 mM MOPS, 15 % isopropanol, pH 7.0
QIAGEN Endofree Maxi QN buffer	1.6 M NaCl, 50 mM MOPS, 15 % isopropanol
QIAGEN QIAquick PB buffer	(proprietary): contains guanidinium hydrochloride, isopropanol
QIAGEN QIAquick PE buffer	(proprietary): contains approx. 80 % ethanol
QIAGEN QIAquick QG buffer	(proprietary): contains guanidinium thiocyanate
Thermo Scientific FastDigest (FD) buffer	(proprietary)
Promega SV Wizard Membrane Binding Solution	4.5 M guanidinium isothiocyanate, 0.5 M KAc, pH 5.0
Promega SV Wizard Membrane Wash Solution	10 mM KAc, 80 % ethanol, 16.7 µM EDTA, pH 5.0
Illumina HT1 buffer	(proprietary): contains: 1 – 10 % NaCl
Illumina PR2 buffer	(proprietary): contains: 1 – 5 % Tris buffer, 1 – 5 % NaCl
PBS (Oxoid)	137 mM NaCl, 3 mM potassium chloride, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.4
PBS/Tween	PBS, 2% Tween-20
Blocking solution	2% BSA (w/v) in PBS, 1 mg/ml heparin

Table 2.2: General ligation setup and conditions using T4 DNA ligase. T4 ligase and buffer, and PEG4000 were purchased from Thermo Scientific.

Component	1X
ddH_2O	Up to 10 µl
10X T4 Ligase buffer	1
PEG4000	1.5
DNA 1	X
DNA 2	Y
T4 DNA ligase (5 Weiss U/µl)	0.5
	10 µl

For Heptad ligations to generate 2.5h and 3.5h variable regions (Figure 2.3), 0.5 µl of High Concentration T4 DNA ligase (Thermo Scientific) was used to improve ligation efficiency. 57 pmol of dephosphorylated Heptad 1 (DNA 2) was ligated with 11.4 pmol of Heptad 2 (DNA 1). 72.4 pmol of dephosphorylated Heptad 1-2 (DNA 2) was then ligated with 24.6 pmol of Heptad 4.5 (DNA 1) using ligation conditions where PEG4000 (Thermo Scientific) was increased to 2 µl per 10 µl ligation. For 3.5h, PEG4000 was also used at 2 µl per 10 µl ligation of 26 pmol of dephosphorylated Heptad 1-2 (DNA 2) with 5.2 pmol of Heptad 2 (DNA 1), and 52.1 pmol of dephosphorylated Heptad 1-2-2 (DNA 2) with 17.4 pmol of Heptad 4.5 (DNA 1).

CIS display sequences were attached to 2.5h and 3.5h variable regions (Figure 2.3) via 'digestion-ligation', performed as detailed in Table 2.3 at 25 °C for 30 min and then 37 °C for 10 min. 48 pmol of MluI-digested (MluI)link-RepA-CIS-Ori was ligated to 40 pmol of AscI-digested Heptad 1-2-4.5 (2.5h) or Heptad 1-2-2-4.5 (3.5h) in a reaction that included MluI and AscI to re-digest any RepA-RepA, 2.5h-2.5h or 3.5h-3.5h that formed from self-ligation at palindromic MluI and AscI sites.

Table 2.3: 'Digestion-ligation' of the 2.5h variable region (Heptad 1-2-4.5) or 3.5h variable region (Heptad 1-2-2-4.5) with RepA sequences. FastDigest (FD) MluI, FD-AscI, FD buffer, ATP, PEG4000 and High Concentration T4 ligase were purchased from Thermo Scientific.

Component	1X
ddH_2O	Up to 130 µl
10X FastDigest (FD) buffer	13
MluI-digested (MluI)-link-RepA-CIS-Ori (24 pmol)	X
AscI-digested Heptad 1-2-4.5 or Heptad 1-2-2-4.5 (20 pmol)	Y
ATP (100 mM); 1 mM final	1.3
PEG4000	26
FD-MluI	2.5
FD-AscI	2.5
High Concentration T4 DNA ligase (30 Weiss U/µl)	7.5
	130 µl

To clone CIS outputs into PCA vectors, 250 ng of NheI/AscI-digested and dephosphorylated pET28a+ (Novagen, Nottingham, UK) (DNA 2) was ligated with 9.4 – 12 ng of NheI/AscI-digested CIS display R4 recovery PCRs (DNA 1), according to Table 2.2 without PEG4000 for 1.5 hours at 25 °C. ≈10 ng of R4 CIS display PCR amplicons was calculated to be necessary to effectively cover the peptide diversity selected by CIS display after selection round R4 (≥25,000 – 82,000 unique peptides identified by deep sequencing, depending on library – see Chapters 4 and 5). Peptide coverage was calculated using Equation 1 (Denault and Pelletier, 2007),

$$\lambda = n(1 - \frac{1}{n})^m \qquad \textbf{(Equation 1)}$$

where λ is the number of variants missing from coverage of a library with m samples (colonies), and n is the library size. This 10 ng estimate was adjusted to 9.4 – 12 ng (depending on library) for efficient cloning using Equation 2 for 250 ng pET28a+,

$$[\text{Insert}]\ (\text{ng}) = \frac{[\text{vector}]\ (\text{ng}) \times \text{insert size (kb)}}{\text{vector size (kb)}} \times \frac{3}{1} \qquad \textbf{(Equation 2).}$$

2.4.5 PCR amplification

PCR was generally performed using proof-reading Phusion Hot Start II (HSII) polymerase (Thermo Scientific) according to Table 2.4 unless otherwise specified, using a PCT-225 DNA Engine Tetrad thermocycler (MJ Research, Waltham, MA, USA). PCR cycles were generally kept to 15 to minimise mutations and over-representation of library members. PCR primers (Table 2.5) were synthesized by Integrated DNA Technologies (Iowa, USA) unless otherwise stated, and were resuspended in ddH_2O or TE buffer (see Table 2.1) for 100 µM. No Template Controls were performed as for sample PCRs except template was replaced with the same volume of ddH_2O. PCR reactions were analysed by agarose gel electrophoresis (see 2.4.6).

2.4.5.1 ProxiMAX Library Construction PCRs

To create Heptad 1, 2 and 4.5 acceptors for library codon addition (see Appendix Figure 2.2 for acceptor sequences), template was either 10 ng/50 µl PCR of pET33+ (Novagen), pUC19 (Invitrogen) or pET33+ plasmid respectively (see Appendix Figure 2.3 for plasmid maps). Templates were amplified with primers 1 and 2 (Heptad 1), 7/8/9/10 and 11 (Heptad 2), or 15 and 16 (Heptad 4.5) (see Table 2.5 for primer sequences). Where subsequent ligations required a 5’ phosphate group, this was introduced via PCR with primers 1 and 3, or 17 and 16.

Following ligation of each library codon to the Heptad 1, 2 and 4.5 acceptors, PCR amplification of 1 µl of the ligation reactions used one of three sets of primers, designated Cycle 1 (C1), C2 and C3, in the sequence C1-C2-C3-C1-…, to minimise carry-over of amplification products from the previous library codon ligation. Primers used for C1 – C3 of each Heptad cassette are displayed in Table 2.6.

Table 2.4: General PCR setup and conditions using Phusion Hot Start II (HSII) polymerase. Phusion HSII DNA polymerase and 5X HF buffer were purchased from Thermo Scientific.

Component	1X
ddH_2O	Up to 50 µl
5X Phusion HF buffer	10
dNTPs	1
Forward primer (10 µM); 0.5 µM final	2.5
Reverse primer (10 µM); 0.5 µM final	2.5
Phusion HSII DNA polymerase (2 U/µl)	0.5
Template DNA (10 – 100 ng, or 1 µl (≈0.2 pmol) of ligation)	X
	50 µl

PCR Cycles		Conditions
Initial denaturation		98 °C, 45 sec
Denaturation	15X	98 °C, 10 sec
Primer Annealing		2 – 5 °C below primer-template T_m, 30 sec
Extension		72 °C, 30 sec/kb
Final extension		72 °C, 8 min

Table 2.5: Oligonucleotide PCR primers (5'→3') for construction of libraries, deep sequencing, CIS display 'recovery' PCRs, and quantitative real-time RT-PCR. Phosphate groups are indicated by "Phos", biotin groups by "Biot", and a three carbon spacer by "SpC3". **Table continued overleaf.**

Primer number and description	Sequence	Primer number and description	Sequence
Heptad 1 Construction		**Heptad 4.5 Construction**	
1, Heptad 1 Acceptor For	GGAAGAGCCTGGTCATCC AGCGGATAG	15, Heptad 4.5 Acceptor For	GGCGCGCCGCTGGTCATCC AGCGGATAG
2, Heptad 1 Acceptor Rev	GCTAGCGGCCATGGTAGA TCCTGTTTCCTGTGGTGG TGTCGATGGTAG	16, Heptad 4.5 Rev	GGAAGAGCGTGGTGGTGTC GATGGTAG
3, Heptad 1 Acceptor Rev (Phos)	Phos-GCTAGCGGCCATGGTAGA TCC	17, Heptad 4.5 Acceptor For(Phos)	Phos-GGCGCGCCGCTGGTCATC
4, Heptad 1 C1 Biotin Rev	Biot-GATCTCACAGTCAGATGG	12, Heptad 2/4.5 C1 Biotin For	Biot-GTGCTACGA TGTCATTGC
5, Heptad 1 C2 Biotin Rev	Biot-TTCGACAGT AGCATCTCG	13, Heptad 2/4.5 C2 Biotin For	Biot-AGGTAGATC AGTGACACG
6, Heptad 1 C3 Biotin Rev	Biot-ATCGACCGT AACTTGAGC	14, Heptad 2/4.5 C3 Biotin For	Biot-ACAGAACAG GCACTTAGG
Heptad 2 Construction		**2.5h and 3.5h Construction and CIS display Sequence Attachment**	
7, Heptad 2 For: library C-terminal E (Glu)	Phos-GAACGTTGGAATCAAGAC GATAG	1, Heptad 1 Acceptor For	GGAAGAGCCTGGTCATCCA GCGGATAG
8, Heptad 2 For: library C-terminal K (Lys)	Phos-AAGCGTTGGAATCAAGAC GATAG	18, Heptad 2 MlyI Biotin Rev	Biot-CTATCGTCTTGAGTCCAAC G
9, Heptad 2 For: library C-terminal Q (Gln)	Phos-CAGCGTTGGAATCAAGAC GATAG	19, Heptad 1 Nested For 1	GGAAGAGCCCAGCGGATAG TTAATGATC
10, Heptad 2 For: library C-terminal R (Arg)	Phos-CGTCGTTGGAATCAAGAC GATAG	20, Heptad 4.5 Nested Rev	GGAAGAGCCGATGGTAGAA CGAAGCG
11, Heptad 2 Rev	SpC3-GGAAGAGCCTGCCGCTTA CCGGATAC	21, CIS display For	CGGCGGTTAGAACGCGGCT ACAATTAATACATAACCCC ATCCCCCTGTTGACAATTA ATCATCGGCTCGTATAATG TGTGGAATTGTGAGCGGAT AACAATTTCACACAGGAAA CAGGATCTACCATGGCC
12, Heptad 2/4.5 C1 Biotin For	Biot-GTGCTACGA TGTCATTGC	22, CIS display Rev	ATGAACGCGGCTACAATTA ATACATAACCTGCATATCT GTCTGTCCACAGG
13, Heptad 2/4.5 C2 Biotin For	Biot-AGGTAGATC AGTGACACG		
14, Heptad 2/4.5 C3 Biotin For	Biot-ACAGAACAG GCACTTAGG		

Table 2.5 continued: Oligonucleotide PCR primers (5'→3') for construction of libraries, deep sequencing, CIS display 'recovery' PCRs, and quantitative real-time RT-PCR. Phosphate groups are indicated by "Phos", biotin groups by "Biot", and a three carbon spacer by "SpC3". **Table continued overleaf.**

Primer number and description	Sequence	Primer number and description	Sequence
12mer, 16mer and 3.4hFosW Construction and CIS display Sequence Attachment		**Heptad Cassette QC Deep Sequencing Sample Preparation PCRs**	
27, 12mer Library For	GGAAACAGGATCTACCATG GCCGCA [NNN] $_{12}$GGCAGC GGTTCTAGTCTAGC	32, Heptad 1 MiSeq Inner For	CTACACGACGCTCTTCCG ATCTNNNNACCACCACAG GAAACAGG
28, 16mer Library For	GGAAACAGGATCTACCATG GCCGCA [NNN] $_{16}$GGCAGC GGTTCTAGTCTAGC	33, Heptad 1 MiSeq Inner Rev	CTGGAGTTCAGACGTGTG CTCTTCCGATCTATCGAC CGTAACTTGAG
29, 3.4hFosW Library For	GGAAACAGGATCTACCATG GCCGCTAGCVWNRMRRMRC TGNHSRMRNHSVWNRMRRM RTTGNNSRMRNHSVWNRMR RMRTTGNHSRMRRMRCTAG AAAAACTGGGCGCGCCTGG CAGCGGTTCTAGTCTAGC	34, Heptad 1/2 MiSeq Outer For	AATGATACGGCGACCACC GAGATCTACACTCTTTCC CTACACGACGCTCTTCCG
26, Link-RepA Rev	CATGATTACGCCAAGCTCA GAA	35, Heptad 1 MiSeq Outer Rev	CAAGCAGAAGACGGCATA CGAGATACATCGGTGACT GGAGTTCAGACGTGTG
21, CIS display For	CGGCGGTTAGAACGCGGCT ACAATTAATACATAACCCC ATCCCCCTGTTGACAATTA ATCATCGGCTCGTATAATG TGTGGAATTGTGAGCGGAT AACAATTTCACACAGGAAA CAGGATCTACCATGGCC	36, Heptad 2 MiSeq Inner For	CTACACGACGCTCTTCCG ATCTNNNNGTGCTACGAT GTCATTGCG
22, CIS display Rev	ATGAACGCGGCTACAATTA ATACATAACCTGCATATCT GTCTGTCCACAGG	37, Heptad 2 MiSeq Inner Rev	CTGGAGTTCAGACGTGTG CTCTTCCGATCTCTTATC CGGTAACTATCGTC
		38, Heptad 2 MiSeq Outer Rev	CAAGCAGAAGACGGCATA CGAGATGATCTGGTGACT GGAGTTCAGACGTGTG
CIS display selection 'recovery' PCRs and Deep Sequencing Sample Preparation PCRs		**CIS→PCA cloning PCRs**	
30, CIS display output Inner For	TCGTCGGCAGCGTCAGATG TGTATAAGAGACAGNNNNC ACACAGGAAACAGGATC	45, CIS-PCA For	CCCCATCCCCCTGTTGAC AATTAATC
31, CIS display output Inner Rev	GTCTCGTGGGCTCGGAGAT GTGTATAAGAGACAGNNNN GTGAAGATCAGTTGGGGC	46, CIS-PCA AscI Rev	AGACTAGAACCGCTGCCC GGCGCGCC

Table 2.5 continued: Oligonucleotide PCR primers (5'→3') for construction of libraries, deep sequencing, CIS display 'recovery' PCRs, and quantitative real-time RT-PCR. Phosphate groups are indicated by "Phos", biotin groups by "Biot", and a three carbon spacer by "SpC3".

Primer number and description	Sequence	Primer number and description	Sequence
2.5h and 3.5h Deep Sequencing Sample Preparation PCRs		**Quantitative real-time RT-PCR**	
39, 2.5h/3.5h Heptad 1 MiSeq Inner For	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNACAGGAAACAGGATCTACC	49, *MMP9* forward	GGCAGCTGGCAGAGGAATAC
40, 2.5h/3.5h Heptad 4.5 MiSeq Inner Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNACTATCCGCTGGATGACC	50, *MMP9* reverse	GGCCCCAGAGATTTCGACTC
41, 2.5h MiSeq Outer For	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG	51, *Cyclin D1* forward	CCCTCGGTGTCCTACTTCAAA
42, 2.5h MiSeq Outer Rev	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC	52, *Cyclin D1* reverse	TCTGTTCCTCGCAGACCTCC
43, 3.5h MiSeq Outer For	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTC	53, *Cyclophilin A* forward	CGCCACCGCCGAGGAAAAC
44, 3.5h MiSeq Outer Rev	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG	54, *Cyclophilin A* reverse	ACCTTGTCTGCAAACAGCTCAAAGG
RepA-CIS-Ori Construction		**Sanger sequencing**	
23, Link-RepA For	GGCAGCGGTTCTAGTCTAGCGGCCCCAACTGATCTTCACCAA	47, 2.5h and 3.5h Sanger sequencing	GAACGCGGCTACAATTAATACATAACC
24, CIS display Extended Rev	CAGGAAACAGCTATGAC	48, DHFR2 Rev (PCA)	GTCTTTCTTCTCGTAGACTTC
25, MluI For	TATTACTACGCGTTGGGCAGCGGTTCTAGTCTAGCGG		
26, Link-RepA Rev	CATGATTACGCCAAGCTCAGAA		

Table 2.6: PCR primers used for amplification of Heptad 1, 2 and 4.5 cassette builds following cycles of library codon ligation. Primer sequences are given in Table 2.5.

Heptad cassette	Cycle of amplification	Primer numbers
1	C1	1 and 4
	C2	1 and 5
	C3	1 and 6
2	C1	12 and 11
	C2	13 and 11
	C3	14 and 11
4.5	C1	12 and 16
	C2	13 and 16
	C3	14 and 16

The 2.5h variable region (Heptad 1-2-4.5) was constructed by two-stage PCR following ligation of Heptad 1 and 2, and ligation of Heptad 4.5 to Heptad 1-2 (see Figure 2.3). The first stage involved amplification of the Heptad 1-2 ligation reaction (0.4 pmol DNA/50 µl PCR) using primers 1 and 18. This was followed by PCR of the Heptad 1-2-4.5 ligation reaction (0.6 pmol DNA/50 µl PCR) using primers 19 and 20.

The 3.5h variable region (Heptad 1-2-2-4.5) was constructed by PCR as for the 2.5h variable region, but with an additional intermediate PCR stage following Heptad 1-2-2 ligation. This PCR amplified 0.7 pmol DNA/50 µl PCR of the Heptad 1-2-2 ligation reaction with primers 19 and 18. The final amplification after Heptad 4.5 ligation was then performed as detailed for 2.5h.

The completed 2.5h variable region was attached to CIS display sequences by a digestion-ligation reaction between AscI-digested 2.5h and MluI-digested (MluI)link-RepA-CIS-Ori constructs (see 2.4.3). This was followed by amplification of 0.5 – 1 pmol DNA/50 µl PCR of this reaction with primers 21 and 22, at 0.4 – 0.5 µM final primer concentration for 20 cycles of PCR.

The completed 3.5h variable region was attached to CIS display sequences in a similar way, except that 0.5 pmol of template was amplified per 50 µl PCR, primers 21 and 22 were used at 0.3 µM final concentration, and 15 PCR cycles were used.

The (MluI)link-RepA-CIS-Ori construct was prepared first by PCR of link-RepA-CIS-Ori according to the conditions in Table 2.7 for Taq polymerase, using 10 ng/50 µl PCR of pCR4-RepA plasmid as template (see Appendix Figure 2.3 for plasmid map) and primers 23 and 24. Following this, 10 ng/50 µl PCR of purified link-RepA-CIS-Ori PCR product was amplified with primers 25 and 26 using the same PCR procedure.

Table 2.7: PCR setup and conditions for Taq polymerase. Taq polymerase, 10X Thermopol buffer, and dNTPs were purchased from NEB (Hitchin, UK).

Component	1X
ddH_2O	Up to 50 µl
10X Thermopol buffer	5
dNTPs	1
Forward primer (10 µM); 0.2 µM final	2.5
Reverse primer (10 µM); 0.2 µM final	2.5
Taq DNA polymerase (5 U/µl)	0.5
Template DNA (10 – 100 ng, or 1 µl (≈0.2 pmol) of ligation)	X
	50 µl

PCR Cycles		Conditions
Initial denaturation		95 °C, 2 min
Denaturation	25X	95°C, 30 sec
Primer Annealing		2 – 5 °C below primer-template T_m, 30 sec
Extension		72 °C, 30 sec/kb
Final extension		72 °C, 7 min

2.4.5.2 12mer, 16mer and 3.4hFosW Library Construction PCRs

The 12mer and 16mer libraries were constructed by PCR using the link-RepA-CIS-Ori construct described in 2.4.5.1 as template DNA and primers 27 or 28 with 26 respectively (see Table 2.5 for primer sequences), following a similar protocol to that in 2.4.5.1. Library forward primers contained NNN degenerate codons (N = A/T/G/C) to encode for library residue options. Finally, CIS display-ready constructs were

generated by PCR using the steps in 2.4.5.1, except using purified 12mer/16mer-link-RepA-CIS-Ori as template (250 ng/50 µl PCR), primers 21 and 22, and 15 PCR cycles.

The 3.4hFosW library was constructed using the same stepwise approach to that for 12mer and 16mer libraries, except that KOD Hot Start polymerase (Novagen) was used instead of Taq polymerase (Table 2.8). To generate 3.4hFosW-link-RepA-CIS-Ori, link-RepA-CIS-Ori was amplified with primers 29 and 26, with primer annealing at 55 °C and 25 PCR cycles. Finally, 3.4hFosW-link-RepA-CIS-Ori template (250 ng/50 µl PCR) was amplified with primers 21 and 22, with primer annealing at 55 °C and 20 PCR cycles to add the *tac* promoter and obtain the CIS display-ready construct. Codon options were encoded in the forward primer using the degenerate codons in Figure 2.1/Table 2.5 (synthesized by GeneLink, Hawthorne, NY, USA).

2.4.5.3 CIS display selection 'recovery' PCRs

The DNA associated with peptide binders pulled down on target-coated beads at the end of each CIS display selection round was amplified to generate sufficient material for the next selection round by 'recovery' PCR. KOD Hot Start DNA polymerase (Novagen, Table 2.8) was used to amplify 10 µl of CIS display R1, R2, R3 or R4 selection round eluate DNA as previously described (Odegrip *et al.*, 2004) with minor proprietary modifications.

Table 2.8: PCR setup and conditions for KOD polymerase. KOD polymerase, 10X KOD buffer, KOD dNTPs and KOD $MgSO_4$ were purchased from Novagen.

Component	1X
ddH_2O	Up to 50 µl
KOD buffer (10x)	5
KOD dNTPs (2mM); 0.2 mM final	5
KOD $MgSO_4$ (25 mM); 1.5 mM final	3
Forward primer (10 µM); 0.2 µM final	1
Forward primer (10 µM); 0.2 µM final	1
KOD polymerase (1 U/µl)	0.5
Template DNA	X
	50

PCR Cycles		Conditions
Initial denaturation		95 °C, 2 min
Denaturation	25-35X	95 °C, 30 sec
Primer Annealing		52 °C, 20 sec
Extension		72 °C, 20 sec/kb
Final extension		72 °C, 7 min

2.4.5.4 Deep Sequencing Sample Preparation PCRs

PCR amplification of samples prior to deep sequencing was achieved via two sequential PCRs, except for 12mer and 16mer selection round outputs that were amplified in a single PCR using long primers 30 and 31. In the two-step PCR approach, an 'inner' PCR amplifies the template and adds sequencing barcodes NNNN, whilst the 'outer' PCR attaches adaptor sequences for immobilisation in the microfluidics flow cell (see 2.4.14 for further details). Inner and outer PCRs were performed as detailed for general PCRs in Table 2.4 with the following modifications. Template for

Heptad 1 and 2 cassette QC sequencing inner PCRs was 1 µl (0.2 pmol) of final codon addition ligations, whilst 1 µl (50 ng) of purified 'recovery' PCR amplicon was used as a template for sequencing CIS display selection round outputs. 2 µl of unpurified inner PCRs was then used as template for outer PCRs. Primers for inner and outer PCRs are displayed in Table 2.5 (primers 32 – 44), and were used at a final concentration of 0.2 µM. 5X Phusion GC buffer was used instead of HF buffer, and PCR cycling was: 98 °C for 45 sec; 10 cycles (five cycles for outer PCR) of 98 °C for 20 sec, 55 °C for 30 sec, and 72 °C for 30 sec; and final extension at 72 °C for eight min.

2.4.5.5 CIS→PCA cloning PCRs

PCRs to reintroduce an AscI restriction site into the R4 recovery PCR products from CIS display selected libraries was performed as in Table 2.4, using primers 45 and 46, 100 ng/50 µl PCR of template, GC buffer instead of HF buffer, and 25 PCR cycles. Full coverage of peptides selected by CIS display for transfer to PCA was achieved by using 100 ng of template DNA, as calculated by Equation 1 (see 2.4.3).

2.4.6 Agarose gel electrophoresis

DNA samples were generally mixed with QIAGEN GelPilot 5X (Manchester, UK) or Bioline 5X (Bioline, London, UK) DNA loading buffer to a final concentration of 1X (see Table 2.1), and loaded onto a 5 mm-thick 2 % agarose (w/v) gel made with TAE or TBE buffer (see Table 2.1) and stained either with 1X SYBR Safe or with 1X SYBR Gold (Invitrogen). Gels were run at 80 – 120 V until bands had been sufficiently separated, and imaged with UV irradiation at 300 nm using a GeneGenius gel doc system (Syngene, Frederick, MD, USA). Sample band molecular weights were estimated by comparison against Hyperladder II or V dsDNA ladders (Bioline, London,

UK). Alternatively, codon addition PCRs during Heptad construction were analysed via gel capillary electrophoresis using a high-throughput QIAxcel (QIAGEN) instrument according to manufacturer instructions. Outer PCRs for deep sequencing were run on a 2 % agarose E-Gel EX gel (Invitrogen) and visualised with blue light, and molecular weights were estimated using an E-Gel 1 Kb Plus DNA Ladder.

2.4.7 DNA immobilization by biotin-streptavidin capture

Biotinylated PCR amplicons were diluted appropriately in 1X B&W buffer (Table 2.1) and captured on Dynal M-280 streptavidin-coated beads (Invitrogen) according to manufacturer instructions.

2.4.8 Restriction Digests

Restriction digests were generally performed as described in Table 2.9 at 37 °C for 15 – 60 min. Digests were analysed by agarose gel electrophoresis (see 2.4.6).

For Heptad cassette MlyI digestion, biotinylated DNA bound to streptavidin-coated beads (see 2.4.7) was incubated at 37 °C with intermittent mixing.

To attach CIS display sequences to the 2.5h library, 180 pmol of Heptad 1-2-4.5 was digested with AscI, and 90 pmol of (MluI)link-RepA-CIS-Ori was digested with MluI before digestion-ligation. For 3.5h, 120.4 pmol of Heptad 1-2-2-4.5 was digested with AscI and 80 pmol of (MluI)link-RepA-CIS-Ori was digested with MluI.

To clone CIS outputs into PCA vectors, 120 µg of pET28a+ vector and 10 µg of each AscI-containing CIS display R4 recovery PCR was digested with NheI and AscI according to Table 2.9 for 1.5 hours at 25 °C.

2.4.9 Dephosphorylation

AscI-digested Heptads and NheI/AscI-digested pET28a+ were dephosphorylated using Antarctic Phosphatase (NEB) (Table 2.10) for 1 hour at 37 °C.

Table 2.9: General restriction digest setup. Fast Digest (FD) MlyI, NheI, AscI, MluI and FD buffer were purchased from Thermo Scientific.

Component	1X
ddH_2O	Up to 20 µl
10X FastDigest buffer	2
Ligated Heptads/PCR amplicon (1 pmol/0.2 µg) or Vector (5 µg)	X
FD-MlyI, NheI, AscI or MluI	1
	20 µl

Table 2.10: General dephosphorylation setup and conditions using Antarctic Phosphatase. Antarctic phosphatase and buffer were purchased from NEB.

Component	1X
ddH_2O	Up to 10 µl
10X Antarctic Phosphatase buffer	1
Acceptor/Heptad (2.5 pmol) or Vector (1 – 5 µg)	X
Antarctic Phosphatase (5 U/µl)	1
	10 µl

2.4.10 Plasmid DNA extraction

GeneJET Plasmid Miniprep kit (Fermentas, Loughborough, UK) or Endofree Plasmid Maxi kit (QIAGEN) columns were used to purify PCA plasmids from BL21 Gold cells or from XL-1 Blue cells respectively, according to manufacturer instructions.

Purified pCR4-RepA, pUC19 and pET33+ plasmids were supplied by Isogenica Ltd. (Little Chesterford, UK).

2.4.11 DNA purification

DNA purifications were typically performed using QIAquick columns (QIAGEN) according to the manufacturer's protocol; modifications are highlighted where applicable. For all purifications, elution buffer (EB buffer or ddH_2O) was heated to 50 °C and incubated on column membranes for 3 – 5 min to improve recovered yield.

For ligations of Heptads 1, 2 and 4.5, and digestion-ligation of AscI-digested library variable regions to MluI-digested RepA, 10 µl of 5 M NaAc pH 5.2 was added to samples before application onto the QIAquick columns, and columns were washed twice with Buffer PB and five times with Buffer PE after sample application to maximise sample purity for optimal ligation efficiency.

Gel purification of PCR reactions of Heptad 4.5 C3 ligation, Heptad 1-2 ligation, Heptad 1-2-2 ligation, (MluI)link-RepA-CIS-Ori, link-RepA-CIS-Ori, and AscI digests of Heptad 1-2-4.5 and Heptad 1-2-2-4.5, was performed before their further use. Excised gel bands were incubated in Buffer QG at 25 °C instead of 50 °C, and 30 µl of 5 M NaAc pH 5.2 was added to samples after gel dissolution. Following sample application to the QIAquick column, a wash with Buffer PB was used to remove traces of agarose and Buffer QG, and Buffer PE washing was increased to three times. Outer PCRs were purified from E-gel EX gels before sequencing following the same protocol.

For construction of 12mer, 16mer and 3.4hFosW libraries, purifications of link-RepA-CIS-Ori amplicon were performed using QIAquick columns, otherwise all

purifications (including gel purifications of *tac*-12mer/16mer/3.4hFosW-link-RepA-CIS-Ori CIS display-ready libraries) used SV Wizard (Promega, Southampton, UK) columns according to manufacturer guidelines. SV Wizard columns were also used to purify recovery PCR amplicons during CIS display selections.

2.4.12 DNA desalting by butanol precipitation

To desalt plasmids post-ligation and before electroporation, 40 µl of ddH_2O was added to 10 µl ligations, followed by 500 µl of butan-1-ol, thorough mixing, and centrifugation at 14,104*g* in an Eppendorf MiniSpin microcentrifuge for 30 min at 25 °C. The supernatant was carefully removed, and the DNA pellet air-dried at 37 °C. The pellet was resuspended in 10 µl of ddH_2O and quantified (see 2.4.13).

2.4.13 DNA quantitation

DNA quantitation was generally performed on a Nanodrop 2000 or 8000 spectrophotometer (Thermo Scientific). 2 µl of samples were analysed, and absorbance at 260 nm was measured against the appropriate blank.

For more accurate quantitation of samples for deep sequencing, Quant-iT™ Picogreen dsDNA fluorescent dye (Invitrogen) was used. Samples and lambda phage DNA standards (3 – 1000 ng/ml) were diluted appropriately in TE buffer in duplicate and then mixed with Picogreen reagent at a 1:400 stock dilution. Sample and standard replicates were transferred to a black 384-well plate in four separate wells, and fluorescence emission at 538 nm (excitation 485 nm) was measured on an Envision (Perkin Elmer, London, UK) plate reader. Fluorescence intensities were averaged across the four wells, and then averaged across two biological replicates.

2.4.14 Deep Sequencing

Deep sequencing was used to record thousands of reads through library samples to sequence individual mutants. Deep sequencing using multiplexing to analyse multiple samples simultaneously was performed on a MiSeq instrument (Illumina, Chesterford, UK) as previously described (Bentley *et al.*, 2008) (Figure 2.4) and according to manufacturer instructions. Briefly, DNA samples were quantified using the Picogreen reagent (see 2.4.13), diluted to 2 nM in EB buffer and pooled equally to obtain an equal number of reads for each sample. The sample pool and 2 nM PhiX DNA control were denatured at 25 °C with 0.1 M NaOH for five min before dilution with ice-cold Buffer HT1 to 4 pM. Sample and PhiX 4 pM solutions were mixed 30:70, and 650 µl of the mixture was added to a MiSeq reagent cartridge.

Deep sequencing was used for ProxiMAX Heptad QC, based on comparison of codon incorporation frequencies observed versus those expected. These frequencies were obtained from raw sequencing data using a proprietary modified version of the open source FastQC analysis tool (Andrews, 2016). Deep sequencing also reported DNA sequences of peptides selected by CIS display after selection rounds R3 and R4.

2.4.15 Sanger Sequencing

For 2.5h and 3.5h, Sanger sequencing was used to confirm RepA-CIS-Ori sequences for CIS display. 30 µl of DNA samples in EB buffer or ddH_2O at 50 – 100 ng/µl were mixed with 4 µl of 10 µM primer 47 (Table 2.5) and sent for analysis by Beckman Coulter Genomics (Takeley, UK). For PCA colonies (see 2.7.1), extracted plasmids were sequenced by Eurofins Genomics (Ebersberg, Germany) as above, except using primer 48.

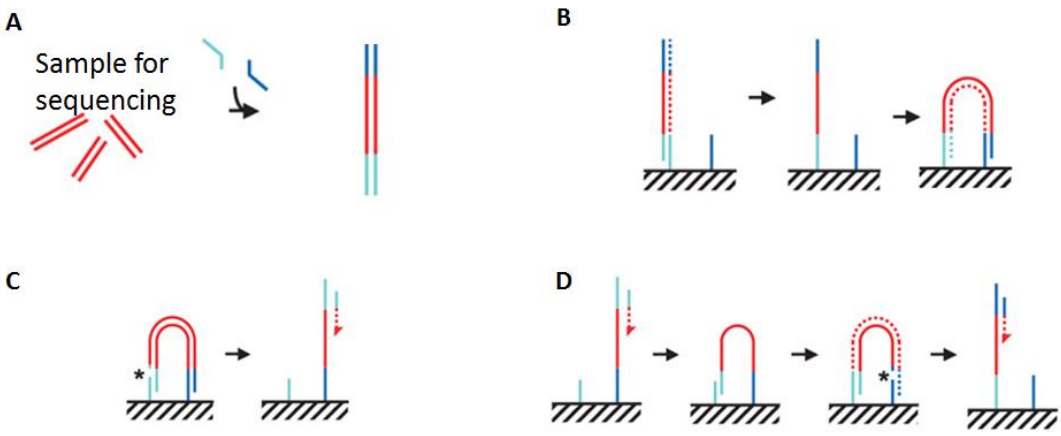


Figure 2.4: Deep sequencing by Illumina MiSeq. A PCR introduces barcodes for cluster identification and adaptor sequences (blue) to either end of samples (red). **B** Sample adaptor sequences anneal to adaptors in a microfluidics flow cell. Immobilised sequences are PCR amplified (dotted line), and the newly synthesized strand remains attached to the flow cell. The 3' adaptor synthesized on the new strand anneals to another flow cell adaptor to create a 'bridge'. This is PCR amplified (dotted line), and the resultant dsDNA is denatured, leaving one strand immobilised and the second strand available to anneal elsewhere on the flow cell (not shown). Repeated cycles generate clusters of copies of each original sample sequence. **C** Single stranded bridges are linearised by cleavage (asterisk). Sequencing from a universal sequencing primer (half arrow) uses fluorescently labelled nucleotides which reversibly terminate sequencing when incorporated. Fluorescent labels are laser-excited, and microscopy optics detect nucleotide fluorescence to assign the sequence read. **D** For paired-end reads (as performed in Chapters 4 and 5), following the first sequencing read (leftmost image), the template strand forms another bridge, and is PCR amplified (dotted line). Single strand cleavage (asterisk) leaves the new strand as a template for the second sequencing reaction which occurs as described above. Paired-end reads increase the length of sequence that can be reliably and accurately sequenced compared to single reads, to sequence entire library variable regions. Adapted from source (Bentley *et al.*, 2008).

2.5 CIS display

CIS display selections were performed as described previously (Odegrip *et al.*, 2004) with minor modifications (stated below). Each selection consisted of four rounds of affinity panning (R1 – R4) of library peptides against biotin-GG-cJun (for

synthesis details, see 2.9.2). Table 2.11 outlines critical parameters and how selection conditions were altered over R1 – R4 to apply selective pressure to library peptides.

Table 2.11: CIS display selection parameters for 12mer, 16mer, 2.5h, 3.5h and 3.4hFosW library selections against cJun. The quantity of cJun available for peptides to bind is reduced from R1 to R2, and from R3 to R4, to provide selective pressure to select the highest affinity peptides. *25 or 12.5 for 12mer/16mer selections †12.5 or 6.25 for 12mer/16mer selections.

	Selection Round			
Parameter	R1	R2	R3	R4
ITT volume (µl)	200	100	100	100
DNA template (µg)	10	5	2.5	2.5
Target (µg)	2 or 0.5	1 or 0.25	1 or 0.25	0.5 or 0.125
Bead volume (µl)	50 or 25	25*	25*	25†
Deselection bead volume (µl)	100	50	50	50
Deselection incubation time (min)	60	60	60	60
Target/beads incubation time (min)	60	60	60	60

First, biotin-GG-cJun peptide was captured on Dynal M-280 streptavidin-coated magnetic beads (Invitrogen) through streptavidin interaction with the peptide's N-terminal biotin tag, according to the manufacturer's instructions. cJun was coated at two different quantities, representing different selection pressures. For R1 of "low stringency" selections (higher target quantity), 2 µg of cJun peptide in PBS was added to each 50 µl of beads, whilst for R1 "high stringency" selections (lower target amount), 0.5 µg of cJun peptide in PBS was added to each 25 µl of beads for R1 (see Chapter 4 for discussion of selection stringency). These coating quantities per µl of beads were optimised in test pulldowns (see Appendix Methods 2.1 for

details and Appendix Figure 2.4 for SDS-PAGE gel data). As detailed in Table 2.11, cJun quantities coated on beads were sequentially decreased at R2, R3 and R4. Coated beads were washed with 5 x 1 ml PBS to remove unbound peptide, and then blocked for one hour in 1 ml blocking solution (Table 2.1) before removal of block from the beads. Separately, 100 µl of M-280 beads per R1 selection were washed and blocked in the same way, and used to remove non-specific binders ("deselection").

Library DNA was expressed using a proprietary *E. coli* cell lysate and buffer (Isogenica Ltd.) with a complete amino acid mixture (Promega) for one hour at 30 °C. These *in* vitro transcription-translation (ITT) reactions were quenched with 600 µl ice-cold blocking solution on ice for 15 min. Blocked ITTs were deselected by adding the blocked, uncoated M-280 beads and rotating for one hour at 25 °C. Deselection beads were carefully removed, and blocked 'high stringency' and 'low stringency' beads were added to deselected ITTs. These selections were mixed in a 2 ml 96-well plate on an automated Kingfisher magnetic particle processor (Thermo Scientific) for one hour at 25 °C, except for 12mer and 16mer selections where this and following steps were completed manually. Selection beads were washed with 4 x PBS/Tween and then 2 x PBS, before manual heat elution of peptide-RepA-DNA complexes at 80 °C for 15 min in 100 µl of 1X Thermopol buffer (NEB). Eluted DNA was amplified by 'recovery' PCR (see 2.4.5.3) to prepare enough material for the next selection round.

By inputting 10 µg of library DNA into the R1 ITTs, the expected number of library peptides presented to cJun in the affinity selections was calculated to be ≈6.1 x 10^{10} peptides. This was based on 10.1 pmol of DNA (≈1500 bp), equivalent to

encoding 6.1 x 10^{12} peptide CIS display constructs, and an expected ITT expression/CIS display coupling efficiency of ≈1 % (Odegrip *et al.*, 2004).

2.6 Microbiology

2.6.1 Media

Media and solutions for bacterial culture are detailed in Table 2.12.

2.6.2 Antibiotics

Antibiotic stocks were prepared as outlined in Table 2.13 and stored at -20 °C.

2.6.3 Bacterial strains

XL-1 Blue and BL21 Gold *E. coli* strains (Stratagene, Stockport, UK; genotypes in Table 2.14) were transformed either with library vectors for amplification prior to starting PCA, or transformed with library and target vectors for the PCA, respectively.

2.6.4 Bacterial Culture

Bacterial cultures were incubated at 37 °C overnight on LB agar, or for up to three weeks on M9 agar, during which time plates were wrapped in Parafilm to prevent them drying out. Liquid cultures were incubated at 37 °C and 220 rpm overnight (LB and 2XYT) or for up to 1 week (M9) in non-baffled flasks.

Table 2.12: Microbiology culture media and solutions compositions. IPTG, LB, 2XYT and bacto-agar were obtained from Melford (Chelsworth, UK), and D-glucose from Fisher (Loughborough, UK).

Medium/Solution	Composition
Luria Bertani (LB) liquid medium	10 g Bacto-trytone, 5 g NaCl, 5 g yeast extract, ddH_2O to 1 l. pH 7.0. Autoclaved.
2XYT liquid medium	16 g Bacto-trytone, 5 g NaCl, 10 g yeast extract, ddH_2O to 1 l. pH 7.0. Autoclaved.
M9 Minimal Media	200 ml 5X M9 Salts, 2 ml 1 M $MgSO_4$, 20 ml 20% D-glucose, 100 µl 1 M $CaCl_2$, ddH_2O to 1 l. pH 7.2. Components 0.22 µ filter sterilised or autoclaved.
5X M9 Salts	200 mM Na_2HPO_4, 100 mM KH_2PO_4, 100 mM NH_4Cl, 50 mM NaCl, ddH_2O to 1 l. Autoclaved.
LB agar	LB broth, 15 g bacto-agar (0.15 % w/v). Autoclaved.
M9 Minimal agar	175 ml M9 agar, 50 ml 5X M9 Salts, 500 µl 1 M $MgSO_4$, 2 ml 50 % glucose, ddH_2O to 175 ml. Components 0.22 µ filter sterilised or autoclaved.
M9 agar	3.7 g bacto-agar (0.5 % w/v), 250 µl 0.1 M $CaCl_2$, ddH_2O to 175 ml. Autoclaved.
100X Electroporation Salts	0.25 M KCl, 1 M $MgCl_2$ in ddH_2O. 0.22 µ filter sterilised.
Isopropyl-β-D-1-thiogalactopyranoside (IPTG)	1 M in ddH_2O. 0.22 µ filter sterilised.

Table 2.13: Antibiotic stock compositions. Ampicillin and kanamycin were purchased from Melford (Chelsworth, UK), and trimethoprim from Duchefa Biosciences (Haarlem, The Netherlands). All antibiotics were 0.22 µ filter sterilised.

Antibiotic	Composition
1000X Ampicillin	100 mg/ml in 70 % ethanol (250 mM)
1000X Kanamycin	50 mg/ml in ddH_2O (100 mM)
500X Trimethoprim	0.5 mg/ml in 70 % methanol (2 mM)

Table 2.14: Genotypes and comments on usage of bacterial strains (Stratagene).

E. Coli strain name	Genotype	Comments relevant to usage
XL-1 Blue (*E. coli* K12 derivative)	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q} *ZΔM15* Tn*10* (*Tet*^{r})]	Endonuclease (*endA* and *hsdR*) deficient for improved quality of extracted DNA, recombination (*recA*) deficient for improved vector insert stability over successive bacterial generations.
BL21 Gold (*E. coli B* derivative)	*F*$^{-}$ *ompT hsdS*(r_B^- m_B^-) *dcm*$^{+}$ *Tet*^{r} *gal endA Hte*	Protein expression strain for PCA assay

2.6.5 Preparation of electrocompetent XL1 Blue and BL21 Gold cells

A 5 ml 2XYT overnight culture was diluted into 500 ml warm 2XYT and cultured until OD_{600} was 0.6 – 0.8 AU. The culture was split into 50 ml Falcon tubes and cooled on ice for 15 min. Subsequent manipulation was performed on ice. Tubes were centrifuged at 2500*g* at 4 °C for 10 min, before washing cell pellets twice with ice-cold ddH_2O (halving total volume to concentrate cells) and centrifuging as described above. Pellets were resuspended in 100 ml total ice-cold 10 % DMSO and centrifuged again, before resuspension in 10 ml total ice-cold 10 % DMSO and re-centrifugation. Finally, cells were resuspended in a total of 1 ml ice-cold 10 % DMSO and OD_{600} of a 1:100 dilution was measured. The cell stock was diluted until the 1:100 dilution OD_{600} measured 0.4 AU, then snap frozen in 80 µl aliquots for -80 °C storage.

2.6.6 Transformation of electrocompetent XL1 Blue and BL21 Gold cells

An 80 µl aliquot of electrocompetent cells were thawed on ice from -80 °C, before addition of 250 ng ice-cold plasmid DNA in ddH_2O, gentle mixing and transfer

to an ice-cold electroporation cuvette (EquiBio, Altrincham, UK). Pulses using settings of 1.7 kV, 400 Ω and 25 µF were applied to cuvettes using a Gene Pulser II instrument (BioRad, Hemel Hempstead, UK). Cells were immediately resuspended in 910 µl warm 2XYT with 10 µl 100X electroporation salts solution (Table 2.12), and were recovered for 75 min at 37 °C and 1300 rpm in a 1.5 ml tube. Cells were plated on LB agar at 1:20, 1:200 and 1:2000 serial dilutions to allow estimation of transformation efficiency by colony counting. The remaining 990 µl of cells was pelleted by centrifugation at 2500*g*, resuspended in 200 µl of the supernatant, and plated.

2.6.7 Bacterial growth quantitation

Bacterial growth was monitored via OD_{600} blanked against the corresponding media (without antibiotics) on an Ultrospec III (Pharmacia LKB Biotech, Uppsala, Sweden) or Jenway 6305 (Staffordshire, UK) UV spectrophotometer.

2.6.8 Cryopreservation of bacterial glycerol stocks

To preserve PCA cultures, an equal volume of liquid cell culture was mixed thoroughly with 80 % autoclaved glycerol (Melford, Chelsworth, UK) to make a 2 ml 40 % glycerol stock, which was snap frozen and stored at -80 °C.

2.7 PCA

PCA selections were performed as described previously (Pelletier *et al.*, 1999) with minor modifications. These consisted of hosting library peptide-mDHFR1 fusions on a pET28a+ plasmid, hosting cJun-mDHFR2 fusions on a pET15b plasmid, and hosting the *lacI*^{q} allele encoding the *lac* repressor for selectively-inducible library and

target expression on the pET15b-cJun-mDHFR1 plasmid instead of on a separate pREP4 plasmid. PCA vector maps are provided in Appendix Figure 2.5.

2.7.1 PCA Single Step selection

Ligation reactions of AscI-containing CIS display R4 recovery PCR amplicons with pET28a+ (see 2.4.3) were butanol precipitated (see 2.4.12) and a total of 1 µg electroporated in multiple transformations (see 2.6.6) into XL-1 Blue cells. Bacterial replication then amplified the library-containing plasmid. Transformants were plated on LB agar containing the appropriate antibiotics (50 µg/ml kanamycin and 100 µg/ml ampicillin). Electroporation and plating of XL-1 Blue transformants was performed until the number of colonies achieved sufficient coverage of the peptides selected by CIS display as estimated using Equation 1 (see 2.4.3) (Denault and Pelletier, 2007), corresponding to 120,000 – 220,000 colonies depending on the library.

Colonies were scraped into 10 ml 2XYT liquid media without antibiotics and recovered for 40 min at 37 °C and 220 rpm. Cells were pelleted by centrifugation in an Eppendorf 5430R centrifuge (1878*g* at 4 °C), and library plasmids were extracted and purified (see 2.4.10). Purified plasmids (1 µg) were electroporated into BL21 Gold electrocompetent cells previously electroporated with pET15b containing cJun-mDHFR1 and the *lac* repressor. Transformants were recovered for 75 min, and then plated to begin the PCA. Cells were plated on:

1. LB agar with 50 µg/ml kanamycin and 100 µg/ml ampicillin at 1:1000 and 1:20,000 dilutions, as positive controls for plasmid transformation efficiency (estimated by counting colonies).

2. M9 agar with 50 µg/ml kanamycin, 100 µg/ml ampicillin and 1 µg/ml trimethoprim, as a negative control for growth without stringent IPTG induction of library peptide and cJun expression.
3. M9 agar with 50 µg/ml kanamycin, 100 µg/ml ampicillin, 1 µg/ml trimethoprim and 1 mM IPTG, as the main PCA plate upon which growing colonies contained peptides that interacted with cJun to reactivate mDHFR.

The number of transformants plated on PCA M9 agar plates was 31,000 – 483,000 depending on the library (see Chapters 4 and 5 for peptide diversity coverage). Colonies that grew on the main M9 agar PCA plate were inoculated into 5 ml LB media overnight cultures with antibiotics, from which plasmids were extracted and library plasmids selectively sequenced by Sanger sequencing (see 2.4.15). In addition, the same colonies were scraped into 5 ml M9 Minimal Media containing antibiotics, trimethoprim and IPTG, and processed to begin PCA Competition Selection passages.

2.7.2 PCA Competition Selection

Competition Selection consists of liquid media co-culturing of bacterial cells containing peptide sequences that were selected in the Single Step selection, with sequential dilution and regrowth (passages "P_0 – P_N"). This applies an even higher selective pressure for cJun interaction and rapid cell growth than Single Step selection, as cells containing peptide sequences directly compete with each other for limited nutrients, to identify the highest affinity peptide(s) for the target.

Scraped and pooled Single Step selection colonies were processed as follows:

1. 1 ml of pooled cells was used to make a 2 ml 40 % glycerol stock (see 2.6.8).

2. 50 µl of pooled cells was diluted into 50 ml of M9 Minimal Media with 50 µg/ml kanamycin, 100 µg/ml ampicillin, and 1 µg/ml trimethoprim, as a negative control for growth.
3. 50 µl of pooled cells was diluted into 50 ml of M9 Minimal Media with 50 µg/ml kanamycin, 100 µg/ml ampicillin, 1 µg/ml trimethoprim and 1 mM IPTG to initiate the next passage growth.
4. 5 µl of pooled cells were plated on LB agar with 50 µg/ml kanamycin and 100 µg/ml ampicillin, three colonies picked, cultured in LB media with antibiotics and Sanger sequenced ("P_0 colonies").
5. And from the remaining pooled cells, plasmids were extracted and Sanger sequenced ("P_0 pool").

Passages were cultured until OD_{600} was ≈0.4 AU before beginning the next passage. Each passage was processed according to the above steps, replacing pooled cells from the Single Step selection for those from the newly grown 50 ml passage. Passaging was typically continued until the pool sequencing returned a single sequence with unchanging base calls, and all colony sequences matched the pool.

2.8 *In silico* peptide sequence analysis

Peptides were analysed using freely available online algorithms to predict biophysical/structural properties and the likelihood of CC formation with cJun.

2.8.1 Agadir

Agadir is an algorithm that predicts helical propensity of an input sequence ((Lacroix *et al.*, 1998) and references therein) by implementation of helix-coil

transition theory (Lifson and Roig, 1961). Peptide sequences (including AS and GAP) were input using the following parameters: N-terminal acetylation, C-terminal amidation, no parameter screening, pH 7, 293 K, ionic strength 0.1 M. For some sequences, prediction at the amino acid level was chosen in the options. Agadir can be found online at http://agadir.crg.es/protected/academic/calculation.jsp.

2.8.2 bCIPA

The bZIP Coiled-coil Interaction Prediction Algorithm (bCIPA) predicts a melting temperature (T_m) for a CC formed from input peptide sequences (Mason *et al.*, 2006). First, an additive stability score is generated from core and electrostatic residue interactions, and averaged residue helical propensities. This score is compared against scores of a training set of 57 dimers whose T_m values have been measured experimentally to estimate an interaction T_m for the input sequences. Peptide sequences (without AS and GAP) were assigned heptad registers beginning with position ***a*** for full length antagonists, or using the heptad register of cJun and aligning shorter/longer peptides to the N- and C-terminus of cJun (both values reported) and truncating cJun/peptide accordingly to consider only the proposed interaction site. bCIPA can be found online at http://www.syntbio.net/bCIPA/.

2.8.3 Base Optimised Weights

Base Optimised Weights (BOW) is an algorithm that works in an analogous way to bCIPA (see 2.8.2) (Fong *et al.*, 2004). BOW uses a base dataset of measured CC stabilities to constrain weighting of possible core and electrostatic interactions between two input sequences in order to obtain a score for the probability of interaction. Peptide sequences (without AS and GAP) were input with their expected

heptad registers (see 2.8.2 for explanation). BOW can be found online at http://compbio.cs.princeton.edu/bzip/.

2.8.4 Prediction of Amyloid Structure Algorithm (PASTA 2.0)

PASTA 2.0 is an algorithm that predicts the amyloidogenic potential of input peptide sequences (Trovato *et al.*, 2006). Predictions are based on the predominance of cross-β-structure within aggregated amyloid fibrils, and the assumption that formation of this structure is identical to β-sheet formation within globular proteins, allowing interaction energies between input peptide residues (an "aggregation energy") to be output. Peptide sequences with AS and GAP were input, and an aggregation energy threshold of >4 kcal/mol applied to outputs, chosen for the best compromise between accuracy (True Positive Rate) and specificity (1 – (False Positive Rate)). PASTA 2.0 can be found online at http://protein.bio.unipd.it/pasta2/.

2.8.5 Zyggregator

Zyggregator is an algorithm analogous to PASTA 2.0 (see 2.8.4) (Pawar *et al.*, 2005). The aggregation propensity of input peptide sequences is calculated as a linear function of weighted parameters for hydrophobicity, charge, hydrophobic-hydrophilic patterning between five adjacent residues, α-helical propensity and β-sheet propensity at a given pH, and is normalised for the aggregation propensity of 100,000 random sequences to give a Z_{agg} score. Input sequences are more aggregation prone than randomly generated sequences if $Z_{agg} > 0$, less prone if $Z_{agg} < 0$, and peptides can be ranked based on increasingly positive Z_{agg} scores as being indicative of greater aggregation potential. Zyggregator can be found online at http://www-mvsoftware.ch.cam.ac.uk/.

2.8.6 DrawCoil

DrawCoil is an algorithm that outputs a helical wheel diagram for a putative coiled coil between two input peptide sequences (Grigoryan, 2016). For all input sequences (without AS and GAP), starting heptad position was assigned as position ***a***, and drawing of potential electrostatic and core-electrostatic interactions was chosen. DrawCoil can be found online at http://www.grigoryanlab.org/drawcoil/.

2.8.7 Logicoil

Logicoil is an algorithm for predicting oligomeric state and helix orientation(s) of a homomeric CC (Vincent *et al.*, 2013). Peptide sequences (without AS and GAP) were input with their expected heptad register (see 2.8.2), using 'Option 1: coiled coil sequence with assigned register for analysis'. Logicoil can be found online at http://coiledcoils.chm.bris.ac.uk/LOGICOIL/.

2.8.8 Peptide sequence alignment using BLASTP

The protein Basic Local Alignment Search Tool (BLASTP) is an algorithm that aligns subject sequences with query (target) sequences, either from the UniProtKB database or input by the user. Alignments are based on similar subsequences, where two residues are similar if they have similar physicochemical properties (Altschul *et al.*, 1990). Peptides without AS and GAP (subject sequences) were aligned with BLASTP against cFos, against FosW, or against Fos homologues (query sequences), or peptides were aligned against query sequence entries in the UniProtKB database. Alignment parameters (recommended by user help documentation) were as follows: maximum target sequences to display was set to 1000; parameter adjustment for

short sequences was opted for; the lower threshold for possible segments was set to three consecutive residue pairs; the similarity matrix selected was PAM-30, the most suitable for input peptide lengths; amino acid compositional adjustments were not applied; and masking for regions of low amino acid compositional complexity was not applied. Other parameters were default settings. BLASTP can be found online at http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins.

2.9 Peptide Characterisation

2.9.1 Materials

For lactam constrained peptide synthesis, Tentagel-S-RAM resin (0.24 mmol loading capacity) was obtained from Sigma-Aldrich (Sydney, Australia), Fmoc-protected L-α-amino acids and Rink Amide MBHA resin were obtained from Novabiochem (Melbourne, Australia), [(6-chlorobenzotriazol-1-yl)oxy-(dimethylamino)methylidene]-dimethylazanium hexafluoro-phosphate (HCTU) was purchased from Chem-Impex (Illinois, USA); and dichloromethane (DCM), dimethylformamide (DMF), diisopropylethylamine (DIPEA), triisopropylsilane (TIPS), piperidine, biochemistry grade trifluoroacetic acid (TFA), diethyl ether and (1H-Benzotriazol-1-yloxy)[tris(dimethylamino)]phosphonium hexafluorophosphate (BOP) were obtained from Auspep (Melbourne, Australia). Acetic anhydride, ninhydrin, HPLC grade acetonitrile and ddH_2O were also obtained from Sigma-Aldrich (Sydney, Australia). Sequences of synthesized peptides can be found in Chapter 3.

For cJun, LIN20 and all other peptide syntheses, H-Rink Amide ChemMatrix™ resin (0.49 mmol loading capacity) was purchased from PCAS Biomatrix Inc. (St. Jean-sur-Richelieu, Canada); Fmoc-protected L-α-amino acids were obtained from

Novabiochem (Watford, UK) or AGTC Bioproducts (Hessle, UK); peptide synthesis grade DMF and diisopropylethylamine (DIPEA) were obtained from AGTC Bioproducts (Hessle, UK); peptide synthesis grade DCM, piperidine and diethyl ether, and biochemistry grade TFA were obtained from Sigma-Aldrich (Gillingham, UK); (Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP) was bought from Novabiochem (Watford, UK); and TIPS was purchased from Acros Organics (Geel, Belgium). For sequences of synthesized peptides, see 2.9.2 below for cJun, Chapter 3 for LIN20, and Chapters 4 and 5 for CIS display and CIS→PCA peptides.

Other reagents were supplied by Sigma-Aldrich (Gillingham, UK) unless otherwise stated.

2.9.2 Peptide Synthesis

Lactam constrained peptides were synthesized as described previously (Rao *et al.*, 2013) by Dr. Timothy A. Hill and Dr. W. Mei Kok of the Institute for Molecular Biosciences, University of Queensland, Australia. Peptides were synthesized by Fmoc SPPS chemistry on a Protein Technologies, Inc. Symphony (Tucson, AZ, USA) at a 0.1 mmol scale using Tentagel-S-RAM resin or Rink Amide MBHA resin. Coupling of non-cyclic amino acids (Table 2.15) was achieved in two identical steps (with a 1 ml DMF wash in between) consisting of 5 equiv. of Fmoc-protected amino acid, 4.5 equiv. of 0.2 M HCTU in DMF as an activator and 10 equiv. of 0.4 M DIPEA in DMF as a base, in a total of 3 ml for 30 min at 25 °C. Following 1 x 5 ml and 3 x 1 ml DMF washes, deprotection was performed with two steps of 1 ml 20 % piperidine in DMF for 3 min. Coupling was then repeated as detailed above after washing with 1 x 5 ml and 4 x 1 ml DMF to remove residual piperidine.

Table 2.15: Identities of Fmoc-protected L-α-amino acids, including side chain protecting groups utilised. *Lys(Mtt) and Asp(OPip) used for lactam cyclisation instead of Lys(Boc) and Asp(OtBu). Amino acids were purchased from Novabiochem (Melbourne, Australia or Watford, UK).

Fmoc-protected amino acids			
Fmoc-Ala-OH	Fmoc-Gln(Trt)-OH	Fmoc-Leu-OH	Fmoc-Ser(tBu)-OH
Fmoc-Arg(Pbf)-OH	Fmoc-Glu(OtBU)-OH	Fmoc-Lys(Boc)-OH	Fmoc-Thr(tBu)-OH
Fmoc-Asn(Trt)-OH	Fmoc-Gly-OH	Fmoc-Met-OH	Fmoc-Trp(Boc)-OH
Fmoc-Asp(OtBu)-OH	Fmoc-His(Trt)-OH	Fmoc-Phe-OH	Fmoc-Tyr(tBu)-OH
Fmoc-Cys(Trt)-OH	Fmoc-Ile-OH	Fmoc-Pro-OH	Fmoc-Val-OH
Fmoc-Lys(Mtt)-OH*	Fmoc-Asp(OPip)-OH*		

Lys and Asp residues to be cyclised were orthogonally protecting using methyl trityl (Mtt) and phenyl isopropyl ester (OPip) protecting groups (Table 2.15), and were coupled for 2 hours. Cyclisation was performed during paused synthesis or post-synthesis, and consisted of removal of Mtt and OPip protecting groups with 5 % TFA in DCM (5 × 2 min) which was monitored via ninhydrin reaction. Washing with 10 ml DMF and DCM removed residual TFA, after which base-catalysed condensation of Lys and Asp side chains to form a peptide bond was achieved by a modified coupling reaction, featuring addition of 1.5 equiv. of PyBOP or BOP, and 2 equiv. of DIPEA, in DMF (1:1) overnight. Cyclised peptide on resin was washed with 10 ml DMF and DCM before checking cyclisation completion using ninhydrin.

Hydrocarbon stapled peptides were also synthesized by Dr. Timothy A. Hill and Dr. W. Mei Kok of the Institute for Molecular Biosciences, University of Queensland, Australia as described previously (Kim *et al.*, 2011).

The CC domain of the cJun peptide used for biophysical studies (sequence: Acetyl-ASIARLEEKVKTLKAQNYELASTANMLREQVAQLGAP-NH_2) (Mason *et al.*, 2006) was bought synthesized and purified to >98 % purity via RP-HPLC by Peptide Protein Research Ltd. (Fareham, UK). AS and GAP additions to the cJun CC represent cloning restriction sites NheI and AscI respectively, which provide good helix-capping for improved helix stability (Doig and Baldwin, 1995). N-acetyl and C-carboxyamide groups on all peptides also aided helix capping (Doig and Baldwin, 1995, Chakrabartty *et al.*, 1993a) and protected peptides from exopeptidase digestion (Sato *et al.*, 2006).

Additional cJun for biophysical characterisations, LIN20 peptide for Chapter 3, and all peptides characterised/used in Chapters 4 and 5 including biotin-GG-cJun (sequence: Biotin-GGASIARLEEKVKTLKAQNYELASTANMLREQVAQLGAP-NH_2) for CIS display selections (see section 2.5), were synthesized in house by the author. Synthesis was performed on a Liberty Blue (CEM; Matthews, NC, USA) microwave synthesizer by Fmoc chemistry on a 0.1 mmol scale using ChemMatrix^{TM} H-Rink Amide resin, and PyBOP/DMF and DIPEA/DMF as activator and base respectively following coupling steps outlined above for non-cyclic amino acids.

As a final step in the synthesis, following Fmoc deprotection all synthesized peptides were N-terminally acetylated. For lactam constrained peptides this was performed by Dr. Timothy A. Hill and Dr. W. Mei Kok of the Institute for Molecular Biosciences, University of Queensland, Australia; for cJun and LIN20 peptides for Chapter 3, and all peptides characterised/used for Chapters 4 and 5 this was performed by the author. Acetylation was achieved using 3 equiv. of acetic anhydride in DMF and 4.5 equiv. of DIPEA in DMF for 20 min at 25 °C. All peptides were then washed on resin with 3 x 5 ml DMF, and air-dried with the aid of 2 x 5 ml DCM to

remove residual DMF. Peptides were cleaved from the resin and their residue side chains deprotected in 10 ml of 95 % TFA, 2.5 % TIPS and 2.5 % ddH_2O for 4.5 hours at 25 °C. Cleavage solution was reduced by evaporation under nitrogen, and resin was removed by filtration. Peptides were precipitated and washed using 4 x 3 volumes of ice-cold diethyl ether, vortexing and centrifugation at 4 °C and 7197*g* in an Eppendorf 5430R. Ether was evaporated off to obtain crude peptides.

2.9.3 Peptide Purification

Crude lactam constrained peptides were purified by Dr. Timothy A. Hill and Dr. W. Mei Kok of the Institute for Molecular Biosciences, University of Queensland, Australia. RP-HPLC was performed on a Shimadzu (Rydalmere, NSW, Australia) LC-20AP Prominence system using a preparative Vydac (Grace, Victoria, Australia) C18 column (10 µm 300 Å, 22 mm x 250 mm), with a gradient from 5 % to 70 % acetonitrile in ddH_2O (both with 0.1 % TFA) over 35 min and at a flow rate of 20 ml/min.

Crude cJun and LIN20 peptides for Chapter 3, and all peptides characterised or used in Chapters 4 and 5 were purified by the author. Semi-preparative RP-HPLC was performed on an Agilent 1100 system (Agilent, Stockport, UK) over a Jupiter Proteo (Phenomenex, Macclesfield, UK) C18 column (4 µm, 90 Å, 10 mm x 250 mm), with a gradient from 5 % to 70 % acetonitrile (VWR, Lutterworth, UK) (both with 0.1 % TFA) over 45 min and at a flow rate of 3.5 ml/min. See Appendix Figure 2.6 for peptide HPLC traces.

2.9.4 Peptide quantitation

Peptide concentrations were determined in ddH_2O, KPP or “low salt” buffer (Table 2.16) against the appropriate blank using the 280 nm absorbance maxima

(absorption coefficient of 1209/M cm) of Tyr residues at solvent-exposed helix positions (Du *et al.*, 1998) on a Varian Cary 50 Conc UV spectrophotometer, except for peptides without Tyr residues which were quantified by dry weight.

Table 2.16: Peptide characterisation buffer compositions. Methyl green was purchased from National Diagnostics (Hessle, UK), β-alanine from Acros Organics (Geel, Belgium), acetic acid from Fisher (Loughborough, UK) and glycerol from Melford (Chelsworth, UK).

Buffer	Components
Potassium phosphate with potassium fluoride ("KPP")	10 mM KH_2PO_4/K_2HPO_4 + 100 mM KF, pH7. 0.22 µ filter sterilised.
Potassium phosphate ("low salt")	10 mM KH_2PO_4/K_2HPO_4, pH7. 0.22 µ filter sterilised.
Native PAGE loading buffer	0.2 % (w/v) methyl green, 375 mM β-alanine acetate pH 3.8, 20% glycerol
Native PAGE running buffer	375 mM β-alanine acetate, pH 3.8

2.9.5 Mass spectrometry

Identities of all peptides post-HPLC purification were confirmed by electrospray mass spectrometry. Lactam constrained peptide +1 charge masses, including correct lactam cyclisations, were confirmed on a Shimadzu LC-MS 2020 instrument (Rydalmere, NSW, Australia) by Dr. Timothy A. Hill and Dr. W. Mei Kok of the Institute for Molecular Biosciences, University of Queensland, Australia.

Identities of cJun and LIN20 for Chapter 3, and all peptides characterised/used in Chapters 4 and 5 were confirmed by the author on a Bruker Daltonics (Coventry, UK) micrOTOF MS using sodium formate clusters generated from NaOH in isopropanol:ddH_2O + 0.1% formic acid for calibration. 10 µl of HPLC fractions were injected in 50:50 ddH_2O:acetonitrile + 0.1 % formic acid into an Agilent 1200 LC system using 70 % methanol at 0.4 ml/min as a mobile phase before electrospray ionisation and detection. Identities were confirmed from +3, +4 and +5 charged ions

(at least two of the three) in the m/z range 500 – 1500. Additionally, MALDI-TOF analysis confirmed the mass of the +1 ion of cJun used for biophysical assays using a Bruker Autoflex Speed MS running flexControl software, and was performed by Mervyn Lewis at the Chemical Characterisation and Analysis Facility, University of Bath, UK. Peptide in aqueous PEG4000 calibrant was mixed with an equal volume of saturated 3,5-dimethoxy-4-hydroxycinnamic acid or 2,5-dihydroxybenzoic acid matrix solution for analysis. See Appendix Figure 2.7 for MS data.

2.9.6 Circular Dichroism (CD)

Coiled coil formation and stability was analysed using an Applied Photophysics (Leatherhead, UK) Chirascan instrument calibrated with (+)-10-camphorsulfonic acid, recording raw ellipticity (mdeg) either of homotypic samples (peptide alone) or heterotypic samples (1:1 peptide:cJun stoichiometric mixture), with a 1 mm pathlength cuvette (Hellma Analytics, Essex UK) and 2 nm bandwidth. All samples were of a total peptide concentration (P_t) of 150 µM unless otherwise stated, and were dissolved in KPP or "low salt" buffer without potassium fluoride (both pH 7). Raw ellipticity was converted to mean residue ellipticity (MRE; deg cm^2/dmol) to normalise for different peptide lengths and concentrations using Equation 3,

$$MRE = \frac{\theta}{10\ l\ Nr\ P_t} \qquad \textbf{(Equation 3)}$$

where MRE is mean residue ellipticity (deg cm^2/dmol), θ is raw ellipticity (mdeg), l is pathlength (0.1 cm), Nr is average number of residues, and P_t is total peptide concentration (M).

2.9.6.1 CD scans

Samples were scanned from 190 – 300 nm in 1 nm increments at 20 °C. For subsequent thermal denaturation (see 2.9.6.2), samples were also scanned at -8 °C, thermally denatured, and then re-scanned at 20 °C to assess denaturation reversibility (in all cases ≥92 %, and typically ≥95 %). Fractional helicity was calculated according to Equation 4,

$$f\mathrm{H} = (\theta_{222} - \theta_c)/(\theta_{222\infty} - \theta_c) \qquad \textbf{(Equation 4)}$$

where $\theta_{222\infty} = (-44000 + 250T)(1 - k/Nr)$ and $\theta_c = 2220 - 53T$. In these equations, θ_{222} is the MRE (Equation 3) from ellipticity measured at 222 nm, $\theta_{222\infty}$ is the expected θ_{222} for a 100 % helical peptide, θ_c is the expected θ_{222} for a random coil, k = 2.4, N_r is the average number of residues in the interacting complex, and T is temperature (293 K) (Rohl and Baldwin, 1997, Luo and Baldwin, 1997).

2.9.6.2 CD Thermal Denaturation

Ellipticity at 222 nm was recorded from -8 °C to 96 °C at 1 °C increments with a tolerance of ± 0.1 °C, allowing 30 seconds equilibration at each temperature, and averaging ellipticity over five seconds. Thermal denaturation data was fit to a two-state model (Equation 5) typical for CC folding and suggested from the cooperative nature of presented denaturation curves,

$$N_2 \rightleftharpoons 2U \qquad \textbf{(Equation 5)}$$

where N_2 is the folded, native (N) CC dimer, and U is the unfolded (random coil) peptide monomer ((Mason *et al.*, 2007a) and references therein).

MRE and derivative F_u data (Equation 8) were fit to a global model by non-linear least-squares regression using Grafit software (Erithracus Software) described

by Equations 6 – 11 to output melting temperature (T_m) values for each CD trace. Linear regression of upper and lower baselines of MRE denaturation data (Equation 3) was defined according to Equations 6 and 7 respectively,

$$ub = ui + usT \qquad \textbf{(Equation 6)}$$

$$lb = li + lsT \qquad \textbf{(Equation 7)}$$

where ub and lb are upper and lower baseline, ui and li are upper and lower y-axis intercept, us and ls are upper and lower slope respectively, and T is temperature (K).

Fraction unfolded (F_u) at each temperature was calculated using Equation 8,

$$F_u = \frac{\theta_{222} - lb}{ub - lb} \qquad \textbf{(Equation 8)}$$

where θ_{222} is MRE (Equation 3) at a given temperature, lb is MRE of the lower baseline linear fit (Equation 6), and ub is MRE of the upper baseline linear fit (Equation 7).

F_u values from Equation 8 were converted to dissociation constant (K_d) values for all temperature data points using Equation 9,

$$F_u = \frac{-K_d + \sqrt{({K_d}^2 + 8K_d P_t)}}{4P_t} \qquad \textbf{(Equation 9)}$$

where F_u is fraction unfolded (Equation 8), K_d is the dissociation constant (M), and P_t is the total peptide concentration (150 μM). Equation 9 is derived from combination of three equations: $K_d = [U]^2/[N_2]$, $P_t = 2[N_2] + [U]$, and $[U] = F_u P_t$, and allows description of F_u in terms of K_d and P_t only for the global fitting procedure. Equation 9 can be solved for K_d using F_u values for each temperature derived by Equations 6 – 8 (Mason *et al.*, 2007a).

The Gibb's free energy (ΔG) can be derived from this K_d by Equation 10,

$$\Delta G = -RTln(K_d) \qquad \textbf{(Equation 10)}$$

where ΔG is the free energy change (kcal/mol) and R is the ideal gas constant (1.9872 cal/mol K), and K_d is from Equation 9.

To complete the global fitting, ΔG values were fit to the Gibbs-Helmholtz equation in Equation 11,

$$\Delta G = \Delta H - (T_A/T_m) \times (\Delta H + RT_m \ln(P_t)) + \Delta C_p(T_A - T_m - T_A \ln(T_A/T_m))$$

(Equation 11)

where ΔG is from Equation 10, ΔH is the change in enthalpy (kcal/mol), T_A is the reference temperature (K), T_m is the melting temperature (K), and ΔC_p is the change in heat capacity (J/K) (Mason *et al.*, 2007a).

However, for the most accurate determination of ΔG of unfolding from K_d, a subset of F_u data from the midpoint of the denaturation transition (which have the best signal to noise ratio of the data collected) were used to calculate the overall K_d, rearranging Equation 9 to give Equation 12 (Mason *et al.*, 2007a),

$$K_d = \left(\frac{2{F_u}^2 P_t}{1-F_u}\right) \qquad \textbf{(Equation 12).}$$

Finally, linear regression of the natural logarithm of these midpoint K_d values vs. temperature provided a means of determining the ΔG of interaction at the reference temperature of 298 K by linear extrapolation, according to Equation 13,

$$\Delta G = mT + C \qquad \textbf{(Equation 13)}$$

where m is the slope of the plot of ln(K_d) vs. temperature (ΔH/R), T is 298 K, and C is the intercept of the plot (-ΔS/R) ((Mason *et al.*, 2007a) and references therein).

2.9.7 Isothermal titration calorimetry (ITC)

Coiled coil interaction thermodynamics were assessed using a Microcal VP-ITC instrument (GE Healthcare, Little Chalfont, UK). cJun at 100 µM – 3 mM was injected into peptide samples at 9 µM – 200 µM in KPP or "low salt" buffer (Table 2.16). Concentrations were varied in order to typically achieve c values ($c = K_A[peptide]$ where K_A is association constant) sufficient for accurate ΔH determination (Wiseman *et al.*, 1989). Change in heating power (µcal/sec) with each injection was integrated to generate binding isotherms, from which ΔH, ΔS, ΔG and K_d values were derived by non-linear least-squares fitting of Equation 14 using Origin software (OriginLab, MA, USA) (Wiseman *et al.*, 1989, Crooks *et al.*, 2011). Equation 14 describes the one-site binding model used for data fitting,

$$q(i) = (n\Delta HVP/2)\left\{[1 + (L/nP) + (K_d/nP)] - [[1 + (L/nP) + (K_d/nP)]^2 - (4L/nP)]^{\frac{1}{2}}\right\} \quad \textbf{(Equation 14)}$$

where $q(i)$ is the heat release (kcal/mol) for the ith injection, n is the interaction stoichiometry, V is the sample cell volume (1.46 ml), P is the total cJun concentration in the syringe (µM), L is the total peptide concentration in the cell after each injection (µM), ΔH is the enthalpy change (kcal/mol) and K_d is the dissociation constant (M).

2.9.8 X-ray crystallography

X-ray crystallography was undertaken by Dr. Nathan Zaccai and Prof. R. Leo Brady at the University of Bristol, UK. Conditions for FosW–cJun crystallisation were identified in a JCSG-Plus (Molecular Dimensions, Newmarket, UK) commercially available 96-well plate crystal screen. FosW–cJun formed diffracting crystals at a concentration of 10 mg/ml in 1.1 M sodium malonate dibasic monohydrate, 0.1 M

HEPES pH 7, and 5 % (v/v) Jeffamine® ED-2003. Diffraction data was collected at the Diamond Light Source (Oxford, UK), and data processing was performed using XDS run via xia2 control software (CCP4, Didcot, UK) in 3dii mode. Phasing was determined by molecular replacement using the crystal structure of cFos–cJun complexed with NFAT (PDB ID: 1A02) (Chen *et al.*, 1998), to generate an electron density map from FosW–cJun data to which FosW and cJun were modelled.

2.9.9 Size-exclusion chromatography

To assess the oligomeric states of peptide–cJun complexes, size-exclusion chromatography (SEC) experiments were performed at 25 °C using a Superdex Peptide 10/300 GL column (GE Healthcare) on an AKTA FPLC system (GE Healthcare) by injecting 100 µl of a 150 µM P_t sample in KPP or "low salt" buffer (both pH 7), at a flow rate of 0.5 ml/min of the same buffer. Elution profiles were recorded via A_{280}.

2.9.10 Native polyacrylamide gel electrophoresis

As an alternative to SEC for assessment of peptide–cJun oligomeric states, samples at a P_t of 600 µM were loaded in 10 µl of native gel loading buffer (Table 2.16) onto a 1 mm-thick 7.5 % polyacrylamide gel (acrylamide/bis-acrylamide 30 % solution from National Diagnostics, TEMED and APS from Fisher) containing native gel running buffer (Table 2.16). Gels were run at 4 °C using a Hoefer SE-250 heat-exchanging gel apparatus (Holliston, MA, USA) for ≤6 hours at 80 V. Gels were fixed in 2 % glutaraldehyde (Fisher, Loughborough, UK) for one hour, and stained for ≥one hour in 0.2 % Coomassie Brilliant Blue R-250 (Thermo Scientific), 10 % acetic acid (Fisher) and 50 % methanol (VWR, Lutterworth, UK), before destaining as necessary in the same solvent lacking the dye (Mason *et al.*, 2007b).

2.10. Human Cell assays

All human cell assays were performed by Mr Samuel Perry at the Institute for Molecular Biosciences, University of Queensland, Australia. Unless otherwise stated, reagents were purchased from Invitrogen (Mulgrave, Victoria, Australia) or Sigma-Aldrich (Castle Hill, NSW, Australia). Fetal calf serum (FCS) was purchased from Bovogen (East Keilor, Victoria, Australia), poly-D-lysine-coated 35 mM dishes were obtained from MatTek (Ashland, MA, USA), 7-aminoactinomycin D was obtained from BioLegend (San Diego, CA, USA), and the ISOLATE II RNA mini kit was purchased from Bioline (Alexandria, NSW, Australia).

2.10.1 Cell culture

MCF-7 cells were cultured in MEM medium. Media were supplemented with 10 % FCS, 100 units/ml streptomycin, 100 units/ml penicillin and 0.01 mg/ml recombinant human insulin ("complete media"), and cells were cultured in a 5 % CO_2 humidified atmosphere (approx. 92 % humidity) at 37°C. Adherent cells were passaged as necessary to prevent confluence; cells were washed with PBS, de-adhered from the culture bottle surface with 0.25 % Trypsin-0.5 mM EDTA for 10 min, and 1 ml of cells was re-seeded at 1 x 10^5 cells/ml in fresh media.

2.10.2 Live Cell Confocal Microscopy

To visualise peptide internalisation into human cells and subcellular localisation, MCF-7 cells (5 x 10^4 cells per dish) were seeded overnight into blocked (poly-D-lysine-coated) 35 mM dishes. On the day of data collection, medium was replaced with serum-free MEM containing FITC-labelled peptides at a final concentration of 10 µM and incubated with cells for six hours. Hoechst 33342 dye at

2.5 µg/ml was used to counterstain cell nuclei for 10 min before washing cells twice with PBS, then three times for five min with 1 mg/ml heparin in PBS, to remove peptide non-specifically bound to cell surfaces. Serum- and phenol red- free MEM was added and cells were transferred to a temperature- and atmosphere- controlled microscope stage (37°C and 5 % CO_2) of a Zeiss LSM 710 FCCS confocal microscope (Carl Zeiss, Munich, Germany). Images were captured with an LD C-Apochromat 63x/1.15 W objective.

2.10.3 Fluorescence-Activated Cell Sorting (FACS)

For FACS analysis, 6-well plates were seeded with MCF-7 cells overnight (5 x 10^5 cells/well). On the day of data collection, culture medium was replaced with serum-free MEM containing FITC-labelled peptides at a final concentration of 10 µM and incubated for one and six hours. Cells were washed twice with PBS, three times for five min each with 1 mg/ml heparin in PBS, and then de-adhered with 0.25 % Trypsin-0.5 mM EDTA for 10 min followed by two PBS washes. 7-aminoactinomycin D and trypan blue (160 µg/ml in PBS) were added to de-adhered cells to aid dead cell removal, and a Gallios flow cytometer (Beckman Coulter, Lane Cove, NSW, Australia) was used for cell sorting. Mean fluorescence intensity from fluorescence probes was calculated from single live cells using Kaluza software (Beckman Coulter).

2.10.4 RNA/cDNA preparation and two-step quantitative real-time RT-PCR

To quantify AP-1 transcriptional activity inhibition, *Cyclin D1* and *MMP9* AP-1 target gene expression levels were assessed in breast cancer cells treated with peptides (without fluorophores). MCF-7 cells were seeded into complete medium overnight at a density of 1.25 x 10^5 cells/well in 12-well plates. On the day of data

collection, medium was replaced with serum-free medium containing peptides at 10 µM final concentration and incubated for 14 hours. PMA (final concentration 50 ng/ml) was added for six hours and total cell RNA was purified using an ISOLATE II RNA mini kit according to manufacturer instructions. From extracted RNA, Superscript III reverse transcriptase and oligo(dT)$_{12\text{-}18}$ primer were used to prepare cDNA according to manufacturer instructions. Quantitative real-time PCR was performed on cDNA amplicons using SYBR Green PCR master mix on an Applied Biosystems ViiA 7 Real-Time PCR system (see Table 2.5 for sequences of primers 49 – 54). *MMP9* and *Cyclin D1* gene expression level was calculated and normalised to *Cyclophilin A* expression. Data was further normalised to vehicle (DMSO) control.

2.10.5 Student's T-test statistical analysis of cell assay data

Unpaired (independent samples) Student's T-tests were performed to assess the statistical significance of findings of cell assays according to Equation 15,

$$t = |\bar{x}_1 - \bar{x}_2|/\sqrt{AB} \qquad \textbf{(Equation 15)}$$

where $A = (n_1 + n_2)/n_1 n_2$ and $B = \left((n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2\right)/(n_1 + n_2 - 2)$. In these equations, $\bar{x}_1$ and $\bar{x}_2$ are the independent sample means, n_1 and n_2 are the independent sample sizes, and s_1 and s_2 are the independent sample standard deviations. T values were compared against the tabulated $t_{n_1+n_2-2}$ distribution to derive reported p values.

Chapter 3 – Truncated helix-constrained Fos-based peptides that inhibit oncogenic Activator Protein-1

3.1 Introduction

Antagonism of the PPI between cFos and cJun AP-1 component proteins is attractive for therapy of many human cancers. However, the PPI interfaces involved are difficult targets for conventional small molecule drugs, and so a more attractive, alternative approach is to downsize one of the interacting surfaces to a bioactive peptide capable of functionally antagonising the PPI. Previous peptidic antagonists, however, still lack drug-like attributes relating to bioavailability to make therapeutic administration attractive (see 1.3.2). Chemical modification of peptides to improve drug-likeness is therefore desirable (Craik *et al.*, 2013, Liskamp *et al.*, 2011).

Structural constraint of helical peptides is one such modification strategy that has received widespread interest. Constrained peptides have been generated against a diverse range of PPIs so far, including those involved in cancers (Grigoryev, 2013, Chang *et al.*, 2013), and have generated peptides displaying efficacy in clinical trials (Grigoryev, 2013). Constraints are covalent cyclisations between amino acid side chains that can pre-organise peptides into α-helical conformations to bind their targets with improved affinities (Rao *et al.*, 2013, Bernal *et al.*, 2007). Constraints can promote a protease-resistant helical conformation, and in some cases may improve cellular uptake of peptides (Schafmeister *et al.*, 2000, Bernal *et al.*, 2007, Walensky *et al.*, 2004). A broad range of constraint chemistries and positions have been described to date. Of these, condensation of amine and carboxyl groups of Lys and Asp side chains at *i→i+4* positions to form lactam constraints has recently been

shown to induce the highest α-helicity in pentapeptide motifs amongst popular all-hydrocarbon 'staples', aryl halides, and other approaches (de Araujo *et al.*, 2014).

Further downsizing of AP-1 antagonists for future therapeutic use was a major goal of this study. However, antagonist shortening is likely to result in a loss of binding free energy (affinity) from loss of enthalpic interactions with the target. Furthermore, as helical peptides are shortened, they lose H-bonds between backbone amide and carboxyl groups that stabilise the helical structure, leading to loss of the target binding conformation. However, ***because helix constraints can potentially minimize the entropic penalty for adoption of a target-binding helical conformation (Rao et al., 2013), it was hypothesized they may compensate for the loss of binding enthalpy upon truncation of an existing AP-1 antagonist, to derive a shorter stabilised peptide that retains binding structure and affinity.***

It was thus the aim of this study to truncate and apply helix constraints to high affinity cJun antagonist FosW, to identify a shorter peptide that retained the attractive binding free energy of the starting peptide. FosW is a cFos mutant peptide selected as the highest affinity binder of cJun from a 62,000 member library screened using an intracellular Protein-fragment Complementation Assay (PCA) (Mason *et al.*, 2006). FosW displays a 70,000-fold lower dissociation constant (K_d) for cJun than native cJun partner cFos, meaning FosW can effectively outcompete cFos (and other native cJun partners) to inhibit formation of cJun-containing AP-1 complexes (Mason *et al.*, 2006), making it an excellent starting point for downsizing. So far, there have been few reports of constrained helical peptide modulators of CC PPIs, with a greater reporting of helix–groove PPI modulation (Bird *et al.*, 2014, Edwards *et al.*, 2016).

However, previously lactam helix constraint successfully generated a shorter derivative of cFos antagonist $JunW_{CANDI}$ that maintained the majority of $JunW_{CANDI}$ binding free energy (Rao *et al.*, 2013). This study suggested the potential for this approach to achieve similar truncation and ΔG retention when applied to FosW.

In this Chapter, determination of the crystal structure of the FosW–cJun LZ CC domain confirmed the expected binding mode of FosW and guided placement of lactam constraints between side chains of substituted Lys and Asp residues at *i→i+4* positions. These were placed in 32mer FosW derivatives, and then in N- and C-terminal truncation mutants, and peptides generated were characterised to see if the FosW ΔG was retained. Lactam constraints allowed generation of a peptide one-third shorter than FosW but which retained the majority of binding free energy for cJun due to a reduced entropic penalty to helix pre-organisation for binding. Cellular uptake of this peptide was demonstrated via appendage of cell- and nuclear-penetrating sequences, but these displayed undesirable effects alone. Instead, replacement of lactam constraints with hydrocarbon constraints in a peptide 22 % shorter than FosW resulted in retention of 88 % of FosW ΔG through entropic pre-organisation to generate an attractive nanomolar affinity cJun binder. This peptide was taken up by breast cancer cells, and preliminary data suggest potential *in cellulo* inhibition of AP-1 activity. Thus, this molecule and future derivatives represent progress towards a therapeutic agent for cancers featuring AP-1 dysregulation.

3.1.1 Experimental Approach

The first objective of this Chapter was to determine the crystal structure of the LZ CC domains of the FosW–cJun complex. This was performed in collaboration

with Dr. Nathan Zaccai and Prof. R. Leo Brady at the University of Bristol. Secondly, lactam helix constraints were introduced between ***b*** and ***f*** or ***f*** and ***c*** positions in the heptad repeats of FosW by solid phase peptide synthesis. This was undertaken in collaboration with Prof. David P. Fairlie, and was performed by Dr. Timothy A. Hill and Dr. W. Mei Kok, all of the Institute for Molecular Biosciences, University of Queensland, Australia. Constraint placement was based on the hypothesis that ***b***, ***c*** and ***f*** positions are solvent exposed and minimally involved in CC stabilisation. *In vitro* biophysical characterisation of cJun binding by lactam constrained and truncated peptides was performed using circular dichroism (CD) spectroscopy to assess helicity and CC binding mode, and by isothermal titration calorimetry (ITC) for accurate binding affinity determination and decomposition of binding free energy into enthalpic and entropic contributions. Human cell assays were performed by Mr Samuel Perry at the Institute for Molecular Biosciences, University of Queensland, Australia. Live-cell fluorescence confocal microscopy and fluorescence-activated cell sorting (FACS) were used to assess peptide cellular uptake and subcellular localisation, and quantitative real-time reverse-transcription PCR (qRT-PCR) was used to assess peptide inhibition of AP-1 transcriptional activity.

3.2 Results

3.2.1 FosW–cJun crystal structure confirms a parallel dimeric CC

To guide the design of the downsized FosW-derived peptides, the crystal structure of FosW–cJun was determined to a resolution of 2.3 Å (Figure 3.1 and Appendix Table 3.1). This is the first crystal structure of an AP-1 antagonist in complex with its target to date, and confirmed the expected parallel heterodimeric CC formed

between FosW and cJun. Importantly, this structure supported the hypothesis that residues at ***b***, ***c*** and ***f*** positions have minimal involvement in interhelical interactions and thus can be replaced without significantly affecting the PPI interface, enabling design of helix-constrained FosW derivatives (Figure 3.2) featuring cyclised Lys and Asp residues at these positions. Furthermore, this structure supports previous hypotheses for selection of more optimal residues in FosW for higher affinity interaction with cJun than those in cFos (Mason *et al.*, 2006).

As expected from other CCs (Mason and Arndt, 2004), all 10 ***a*** and ***d*** position residues in FosW–cJun are aligned for burial of their hydrophobic side chains away from solvent, a major enthalpic contribution to dimerisation (Acharya *et al.*, 2006), with interatomic distances in agreement with stabilising interhelical ***a***–***a'*** and ***d***–***d'*** interactions (Glover and Harrison, 1995, O'Shea *et al.*, 1991). Compared to cFos, predominant selection of bulky hydrophobic side chains of L (opposite cJun I or V) and I (opposite cJun V) provides better core packing and solvent shielding than T or K at cFos ***a*** positions, and improved enthalpy to drive dimerisation (Acharya *et al.*, 2006). Furthermore, selection of I rather than V at ***a*** positions is favoured opposite small side chains of V and A due to slight increases in van der Waals' interactions. Finally, opposite N in the cJun core, the N of FosW is much less sterically challenged than the long side chain of K in cFos, and thus is less disrupting (Acharya *et al.*, 2006).

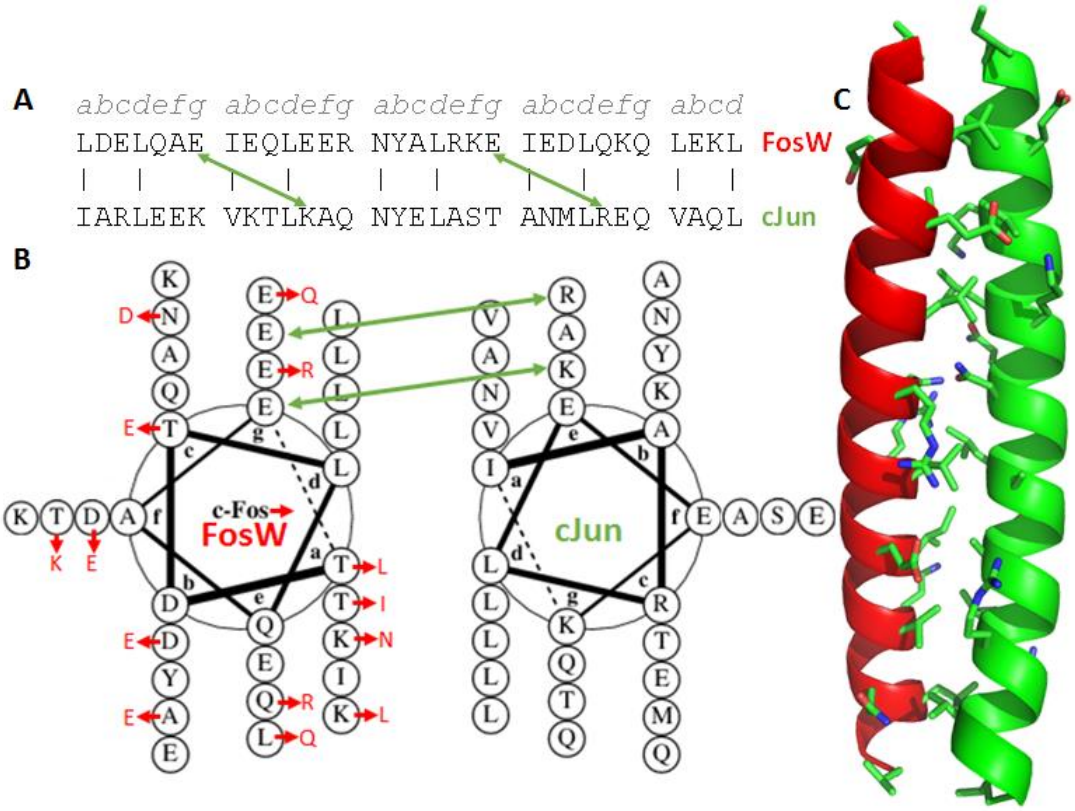


Figure 3.1: Crystal structure of the FosW–cJun coiled coil. A FosW and cJun sequences. Stabilising interhelical hydrophobic interactions ($\boldsymbol{a}_i$-$\boldsymbol{a'}_i$ and $\boldsymbol{d}_i$-$\boldsymbol{d'}_i$) are shown as vertical lines and specificity-conferring interhelical electrostatic interactions ($\boldsymbol{g}_i$-$\boldsymbol{e'}_{i+7}$ and $\boldsymbol{e}_i$-$\boldsymbol{g'}_{i-7}$), are shown as green arrows. Heptad register in italics above FosW sequence. **B** Helical wheel diagram of FosW (showing residues selected from cFos library in red) in interaction with cJun, including interhelical interactions between ***e*** and ***g*** residues (green arrows). Adapted from source (Mason *et al.*, 2006). **C** Crystal structure of the FosW–cJun CC to 2.3 Angstroms. cJun is shown as a green ribbon, FosW as a red ribbon. Side chains for ***a***, ***d***, ***e*** and ***g*** residues only are shown, using CPK colouring. Ribbons go from N-terminus (top) to C-terminus (bottom). See Appendix Table 3.1 for crystallisation and structure solving parameters.

Residues at ***e*** and ***g*** positions mostly behave as hypothesized, with one exception. Of the eight ***e*** and ***g*** positions that are present in FosW, ***e***1 and ***g***4 cannot form interhelical electrostatic $\boldsymbol{g}_i$–$\boldsymbol{e'}_{i+7}$ and $\boldsymbol{e}_i$–$\boldsymbol{g'}_{i-7}$ interactions as their partner residues are not present in the crystallised CC region. Of the remaining six positions, as expected two form potentially stabilising and/or specificity-conferring attractive

interactions, between ***g***1 E and ***e'***2 K and between ***g***3 E and ***e'***4 R (Krylov *et al.*, 1998, Graddis *et al.*, 1993), with interatomic distances in agreement with previous crystal structures (Glover and Harrison, 1995, O'Shea *et al.*, 1991). Residues at positions ***g***2, ***e'***3, ***e***3, ***g'***2, ***e***4 and ***g'***3 could have been selected for similar interhelical interactions, but instead are too distant to form contacts. Finally, the expected ***e***2–***g'***1 E–K interaction is not supported by the crystal structure; instead the ***g'***1 K amine faces away from the ***e***2 E, potentially attracted to the ***g'***2 Q and/or ***c'***2 T, the former of which is free from interaction with ***e***3 R. The ***e***2–***g'***1 interaction was expected to occur between FosW and cJun and stabilise this complex based on its presence in the crystal structure of cFos–cJun (Glover and Harrison, 1995), and the fact that neither residue was randomised in the FosW library selection, and thus its absence will be interesting to investigate further. It was previously unknown why ***g2*** R was selected opposite ***e'3*** A: the crystal structure suggests that ***g***2 R interacts with solvent to aid solubility or potentially is involved in intrahelical attraction with ***g***3 E, though this is unlikely to contribute significantly to helix or CC adoption. The Glu residue at ***e***4 does not interact with ***g'***3 Thr and, as alternatively hypothesized, instead shields FosW ***a***4 I from solvent with its hydrocarbon chain (O'Shea *et al.*, 1991). Similarly, ***e***1 and ***g***4 Q residues, which have no $\boldsymbol{g'}_{i-7}$ or $\boldsymbol{e'}_{i+7}$ partner, shield ***a***2 I and ***d***4 L core residues.

Finally, ***b***, ***c*** and ***f*** position residues appear to perform expected functions. Residues in these positions are thought to be generally involved in solvent interaction for solubility, or form intrahelical interactions which promote the helical conformation in monomeric peptides and consequently in coiled coils (Smith and Scholtz, 1998, O'Shea *et al.*, 1991, Burkhard *et al.*, 2002). In the FosW–cJun crystal structure the majority of these outerface residues appear to interact with solvent, as

previously suggested (Mason *et al.*, 2006). It was previously hypothesized that ***f***2 E and ***c***4 D residues could have been chosen in FosW for attractive intrahelical interactions, however the crystal structure reveals that these side chains are too distant from proposed partner atoms at *i→i+3* or *i→i+4* positions for meaningful interaction. In fact, the only possible intrahelical interactions involve previously overlooked ***b***3 Y, ***f***3 K, ***b***4 E and ***f***4 K residues. It is likely these interactions consist of a single salt bridge between ***f***3 K and the ***b***4 E that faces it, with attraction of ***f***4 K towards the ***b***4 E and a cation–π attraction between ***b***3 Y and ***f***3 K (Andrew *et al.*, 2002, Slutsky and Marsh, 2004). The contribution of these intrahelical interactions to α-helix adoption and dimerisation free energies is expected to be small (Smith and Scholtz, 1998, Slutsky and Marsh, 2004, Andrew *et al.*, 2002), such that their loss is likely to have minimal effect on CC stability.

Thus, this crystal structure evidences that ***b***, ***c*** and ***f*** residues are relatively free from involvement in crucial CC-stabilising interactions, such that replacement of residues at these positions with lactam helix constraints in FosW derivatives should not interfere significantly with CC formation. This crystal structure will also enable design of future cJun antagonists, and further validates the use of *in vivo* library selection systems such as PCA for identifying successful cJun binders.

3.2.2 Downsizing FosW via helix-inducing constraints

Based on confirmation that ***b***, ***c*** and ***f*** residues are minimally involved in interhelical interactions between FosW and cJun, *i→i+4* lactam helix constraints were introduced between ***b*** and ***f*** or ***f*** and ***c*** residues in the heptad repeats of FosW. Initially, single lactam constraints were introduced in a scan across all four heptads

and one half-heptad of FosW (minus AS and GAP helix-capping motifs to begin the truncation process) (Figure 3.2, compounds **1-5**), to identify constraints that were tolerated for retained cJun interaction. Following this, introduction of multiple constraints tested even greater induction of helicity (peptides **6**-**9**). Constraint was then combined with N- and/or C-terminal truncation (peptides **10**-**20**) to identify truncated antagonists with retained binding free energy. Finally, cellular uptake and *in cellulo* AP-1 inhibition of promising truncated peptides was evaluated.

For presented peptide designs, circular dichroism (CD) scans (Figure 3.3) were used to determine α-helicity based on intensity of 222 nm and 208 nm minima, and likelihood of CC formation, where CCs have a 222:208 ratio >≈0.9 (Zhou *et al.*, 1992, Lau *et al.*, 1984). Successful heterodimerisation of peptides with cJun was reported where the heteromeric sample (1:1 stoichiometric mix of peptide and cJun) exceeded the calculated average of cJun alone and peptide alone traces indicative of non-interaction. CD thermal denaturations were fit to a 2-state unfolding model to estimate melting temperature (T_m) values for CC dissociation ((Mason *et al.*, 2007a) and references therein), where again heterodimerisation was reported for traces exceeding the average helicity and/or T_m (Figure 3.4, Table 3.1). For CD analyses, an internal Tyr residue is known to contribute to helicity (Chakrabartty *et al.*, 1993b), but inclusion in virtually all peptides in this study allowed comparison across peptides. Isothermal titration calorimetry (ITC) isotherms were also fit to a 2-state binding model and were used to decompose binding enthalpy (ΔH) and entropy (TΔS), and quantify retention of binding free energy (ΔG) (Wiseman *et al.*, 1989) (Figure 3.5 and Table 3.1).

	Heptad 1	Heptad 2	Heptad 3	Heptad 4	Heptad 4.5	Truncation vs. FosW	Peptide–cJun $K_{d(20°C)}$ (nM) (ITC)	Peptide–cJun Helicity (%) (CD)
	a b c d e f g	*a b c d e f g*	*a b c d e f g*	*a b c d e f g*	*a b c d*			
FosW:	AS**LDELQAE**	**IEQLEER**	**NYALRKE**	**IEDLQKQ**	**LEKL**GAP	N/A	39 ± 11	37
1	Ac-LDELQAE	IEQLEER	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ2 CΔ3	9.6 ± 5.7	44
2	Ac-LDELQAE	IEQLEER	NYALRKE	IKDLQDQ	LEKL-NH_2	NΔ2 CΔ3	16 ± 4.7	62
3	Ac-LDELQAE	IEQLEER	NKALRDE	IEDLQKQ	LEKL-NH_2	NΔ2 CΔ3	ND	29
4	Ac-LDELQAE	IKQLEDR	NYALRKE	IEDLQKQ	LEKL-NH_2	NΔ2 CΔ3	40 ± 11	60
5	Ac-LKELQDE	IEQLEER	NYALRKE	IEDLQKQ	LEKL-NH_2	NΔ2 CΔ3	29 ± 19	74
6	Ac-LDELQAE	IEQLEER	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ2 CΔ3	19 ± 6.6	56
7	Ac-LDELQAE	IEQLEKR	NYDLRKE	IEDLQKQ	LEDL-NH_2	NΔ2 CΔ3	8.6 ± 1.8	73
8	Ac-LDELQKE	IEDLEER	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ2 CΔ3	9.9 ± 4.0	75
9	Ac-LKELQDE	IEQLEER	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ2 CΔ3	11 ± 5.5	77
10		Ac-LEKR	NYDLRKE	IEDLQKQ	LEDL-NH_2	NΔ12 CΔ3	ND	34
11		Ac-LEKR	NYDLRKE	IEDLQKQ	LEDL-NH_2	NΔ12 CΔ3	ND	14
12	Ac-LQKE	IEDLEER	NYALRKE	IEDL-NH_2		NΔ5 CΔ10	ND	25
13	Ac-LQKE	IEDLEKR	NYDLRKE	IEDL-NH_2		NΔ5 CΔ10	ND	27
14	Ac-LKELQDE	IEQLEER	NYALRKE	IEDL-NH_2		NΔ2 CΔ10	2200 ± 450	44
15	Ac-LKELQDE	IEQLEKR	NYDLRKE	IEDL-NH_2		NΔ2 CΔ10	1000 ± 560	62
16	Ac-LDELQAE	IEQLEKR	NYDLRKE	IEDL-NH_2		NΔ2 CΔ10	ND	49
17		Ac-IEQLEKR	NYDLRKE	IEDLQKQ	L-NH_2	NΔ9 CΔ6	ND	44
18	Ac-LKAE	IDQLEER	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ5 CΔ3	56 ± 14	69
19	Ac-LQAE	IKQLEDR	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ5 CΔ3	110 ± 50	60
20		Ac-IKQLEDR	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ9 CΔ3	2000 ± 420	62
LIN20		Ac-IKQLEDR	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ9 CΔ3	55,000 ± 9700	39
LIN20-TAT		Ac-IKQLEDR	NYALRKE	IEDLQKQ	LEDL RKKRRQRRR-NH_2	NΔ9 CΔ3	ND	59
Pal-**20**		Pal-IKQLEDR	NYALRKE	IEDLQKQ	LEDL-NH_2	NΔ9 CΔ3	ND	28
20-TAT		Ac-IKQLEDR	NYALRKE	IEDLQKQ	LEDL RKKRRQRRR-NH_2	NΔ9 CΔ3	5600 ± 380	70
20-NLS-TAT		Ac-IKQLEDR	NYALRKE	IEDLQKQ	LEDL PKKKRKVYGRKKRRQRRR-NH_2	NΔ9 CΔ3	7600 ± 460	39
NLS-TAT					Ac-PKKKRKVYGRKKRRQRRR-NH_2	N/A	NF	11
18HC	Ac-LXAE	IXQLEER	NYALRKE	IEDLQXQ	LEXL-NH_2	NΔ9 CΔ3	320 ± 50	44
20HC		Ac-IXQLEXR	NYALRKE	IEDLQXQ	LEXL-NH_2	NΔ9 CΔ3	15,000 ± 2400	35

Figure 3.2: Helix-constrained and truncated FosW derivatives. Design template FosW sequence in bold. Heptad register ***a*** **–** ***g*** through each heptad of FosW in italics. Helix constraints between Lysine and Aspartate are shown in blue. Standard single letter amino acid code except Pal = palmitic acid, TAT = Tat peptide of HIV-1 (Debaisieux *et al.*, 2012), NLS = the monopartite SV40 large T-antigen nuclear localisation signal (Lanford *et al.*, 1986), Ac- = N-terminal acetyl modification, -NH_2 = C-terminal amine, X = (S)α-(2'-pentenyl)alanine. Dissociation constants (K_d) measured by isothermal titration calorimetry (ITC), and fractional helicities calculated from ellipticity measured by circular dichroism spectroscopy (CD) using Equation 4 (see 2.9.6.1). Values to 2 s.f..

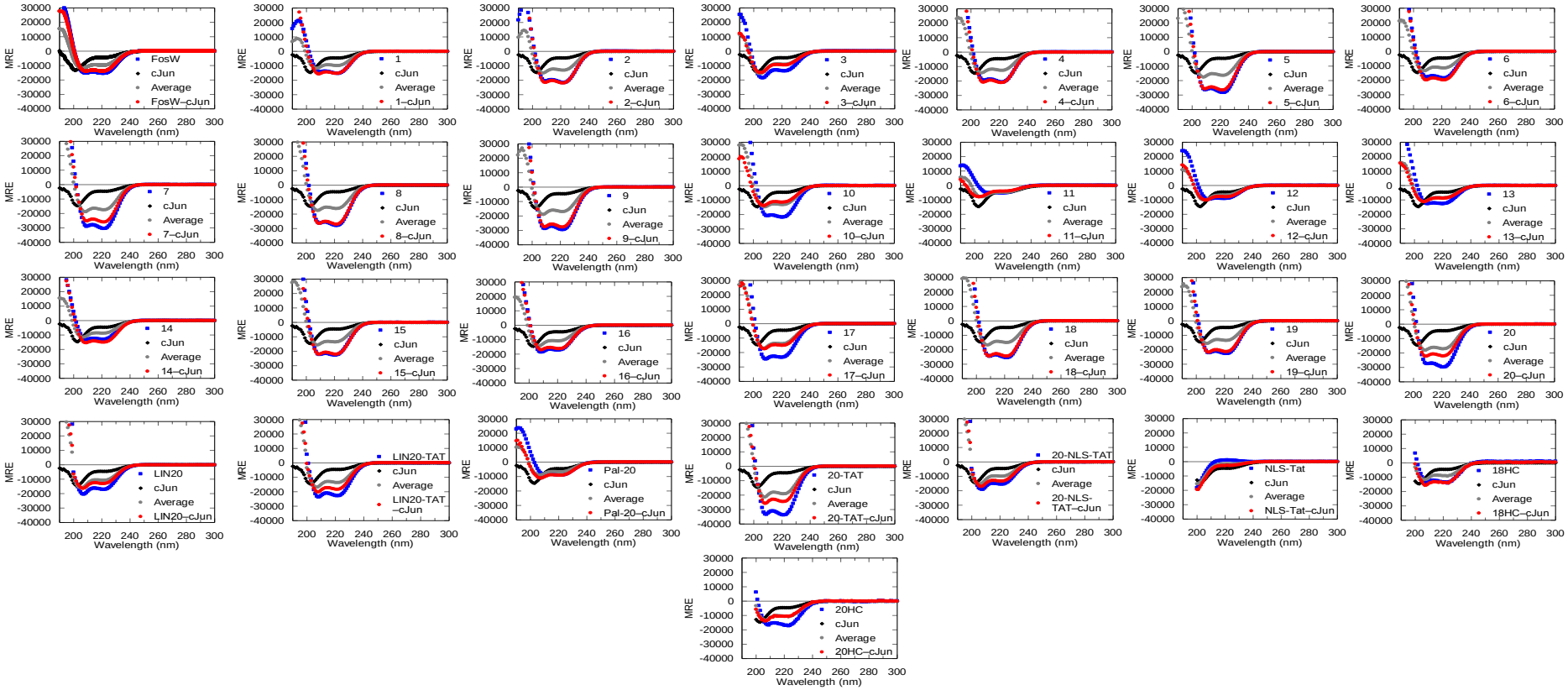


Figure 3.3: Helix constrained FosW derivatives form highly helical CCs with cJun despite truncation. CD scans at 20 °C. Data reported as change in mean residue ellipticity (MRE; units deg cm^2 dmol^{-1}, to allow for comparison between peptides of different lengths), as a function of CD ellipticity over the wavelength range 190 nm (or 200 nm) – 300 nm. Blue squares: peptide alone, red circles: equimolar mix of peptide with cJun, grey diamonds: average of cJun and peptide alone, black triangles: cJun alone. Peptides interact with cJun where MREs for peptide:cJun mixes exceed that of the calculated average trace of peptide and cJun traces (which represents non-interaction). MRE was calculated according to Equation 3 (see 2.9.6.1). FosW, cJun and FosW–cJun traces from Mason *et al.* (2006) are in excellent agreement with that reported here.

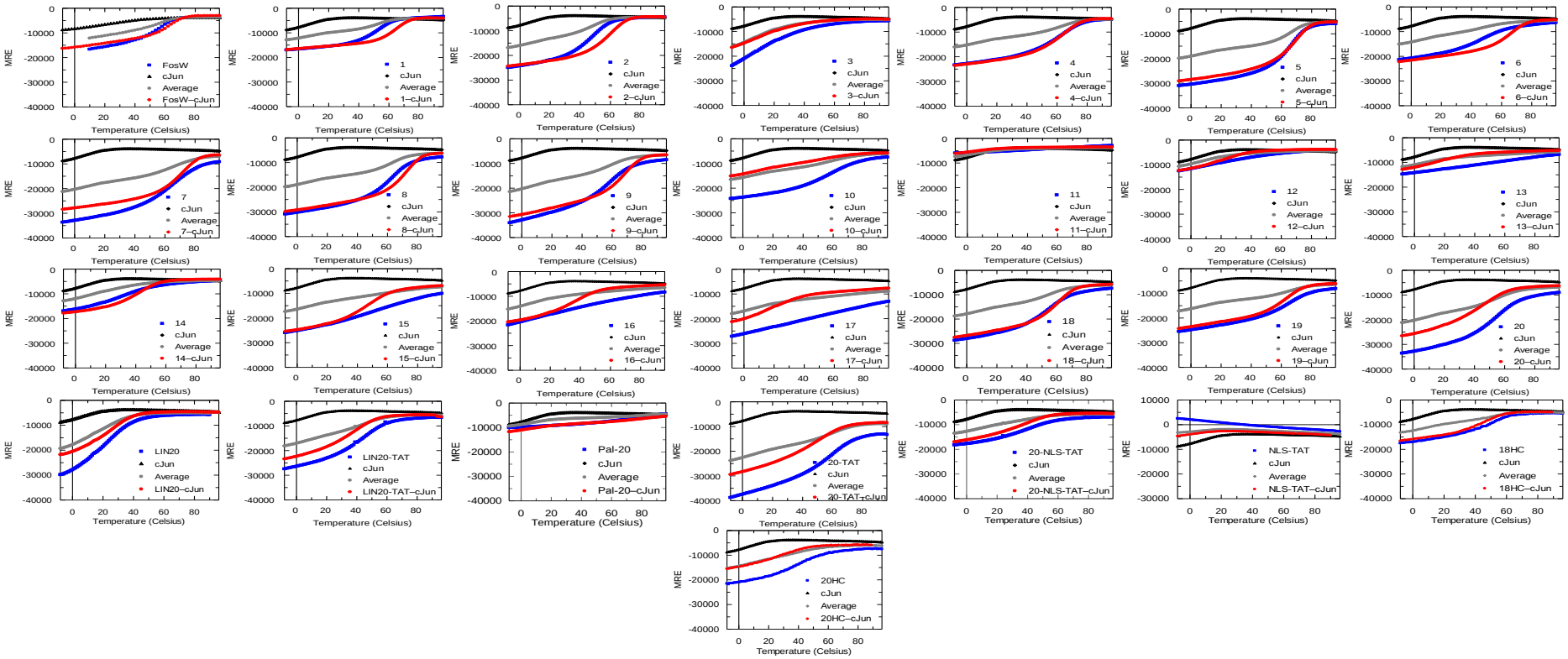


Figure 3.4: Helix constrained FosW derivatives form CCs with cJun resistant to thermal denaturation despite truncation. CD thermal denaturation profiles. Data reported as change in mean residue ellipticity (MRE; units deg cm^2 $dmol^{-1}$, to allow for comparison between peptides of different lengths), as a function of CD ellipticity at 222 nm with temperature. Blue squares: peptide alone, red circles: equimolar mix of peptide with cJun, grey diamonds: average of cJun and peptide alone, black triangles: cJun alone. Peptides interact with cJun where MRE and/or T_m for peptide:cJun mixes exceed that of the calculated average trace of peptide and cJun traces (which represents non-interaction). MRE was calculated using Equation 3 (see 2.9.6.1), and all traces were fit to Equations 6 – 13 (see 2.9.6.2) where lower baselines and sigmoidal traces were present (data in Table 3.1). FosW, cJun and FosW–cJun traces from Mason *et al.* (2006) are in excellent agreement with that reported here.

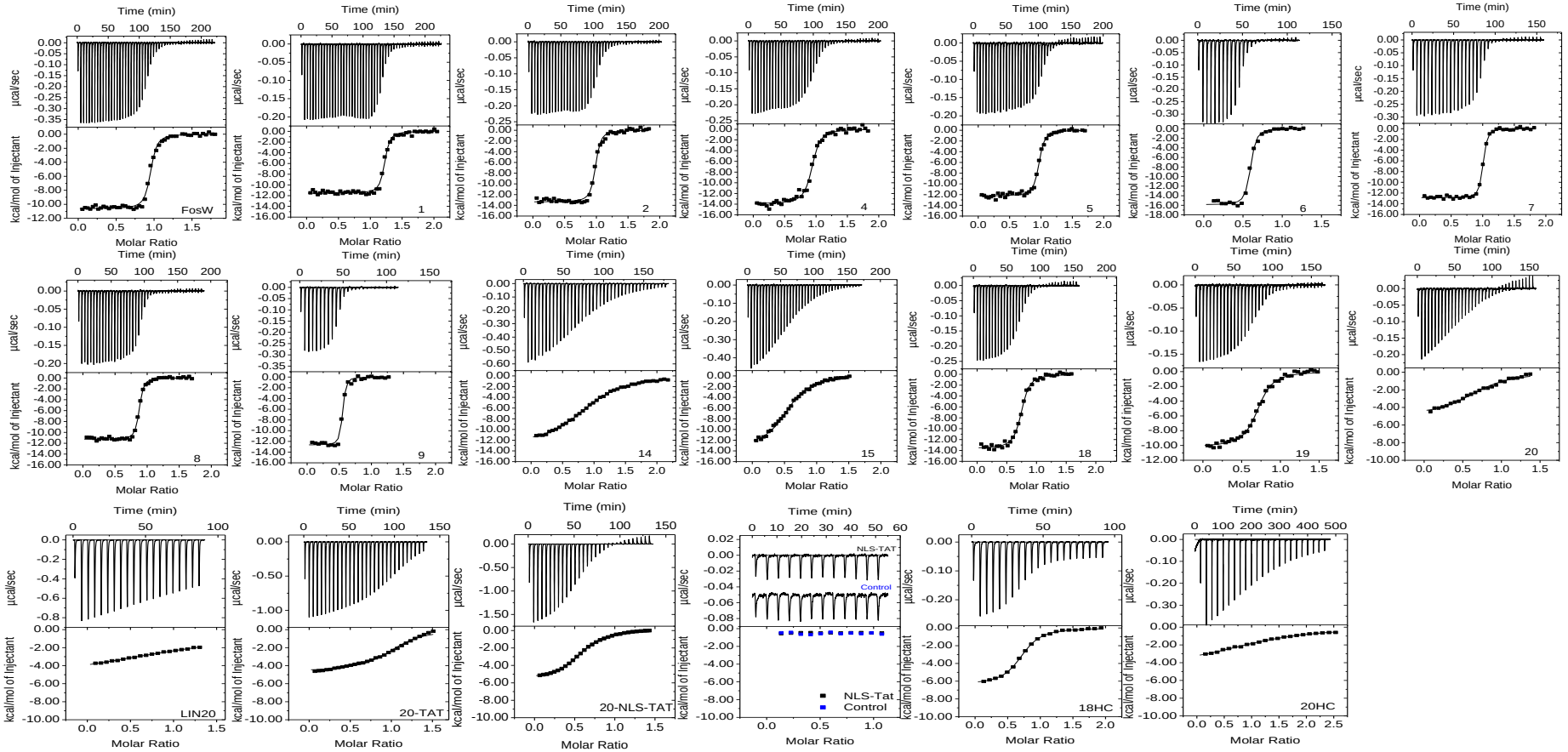


Figure 3.5: Helix constrained FosW derivatives bind cJun with good retention of binding free energy despite truncation. Isothermal titration calorimetry raw isotherms (top panels) and fitted data (bottom panels) (both baseline corrected) for select peptides with cJun. On the fitted data plots, solid lines represent fits generated by non-linear least-squares fitting to Equation 14 (see 2.9.7). Raw isotherm injections for the buffer into buffer control (blue text label) on the NLS-TAT plot are translated by -0.05 µcal/sec for clarity relative to those for cJun into NLS-TAT.

Table 3.1: Thermodynamic parameters for cJun interaction with lactam constrained FosW derivative peptides. Data from CD and ITC measurements (2 s.f.). θ is raw CD ellipticity (mdeg). Fractional helicity is as calculated using Equation 4 (see 2.9.6.1). TΔS is calculated as ΔH – ΔG from the Gibbs-Helmholtz equation (Equation 11, see 2.9.6.2). FosW–cJun CD values from Worrall and Mason (2011). CD values are taken from representative single measurements, which are typically reproducible in biological replicates to ± 1 °C for T_m, within 5 % for fractional helicity and 222:208 ratio ($\theta_{222}/_{208}$), and within 10 % for K_d and ΔG (data not shown). ITC values are the arithmetic mean of at least two independent titrations ± SDs, except values from Worrall and Mason (2011) indicated with an asterisk and values for **20**-TAT and **20**-NLS-TAT (single titrations and fitting errors). CD and ITC data generally agree to within 15 % for ΔG and an order of magnitude for K_d. "ND" = not determined, "NF" = not fit (CD traces lacking lower baseline, ITC fits poor).

Complex	T_m (°C) (CD)	Peptide–cJun Fractional Helicity (%) (CD)	$K_{d(20°C)}$ (nM) (CD)	$\Delta G_{(20°C)}$ (kcal/mol) (CD)	$K_{d(20°C)}$ (nM) (ITC)	N interaction stoichiometry (ITC)	$\Delta G_{(20°C)}$ (kcal/mol) (ITC)	$\Delta H_{(20°C)}$ (kcal/mol) (ITC)	$T\Delta S_{(20°C)}$ (kcal/mol) (ITC)	Peptide–cJun $\Theta_{222}/_{208}$ (20°C) (CD)
cFos–cJun*	16	28	320,000	-5.5	27,000	1.1 ± 0.01	-6.1 ± 0.39	-0.82 ± 0.36	5.3 ± 0.53	0.75
FosW–cJun	63	37	4.0	-11	39 ± 11	0.99 ± 0.08	-9.9 ± 0.16	-10 ± 0.42	-0.46 ± 0.46	1.0
1–cJun	65	44	0.51	-13	9.6 ± 5.7	0.92 ± 0.42	-11 ± 0.35	-12 ± 0.62	-1.1 ± 0.67	0.98
2–cJun	60	62	19	-10	16 ± 4.7	0.95 ± 0.03	-10 ± 0.17	-13 ± 0.30	-3.0 ± 0.41	1.0
3–cJun	NF	29	NF	NF	ND	ND	ND	ND	ND	0.70
4–cJun	63	60	33	-10	40 ± 11	0.86 ± 0.10	-9.9 ± 0.16	-14 ± 0.33	-3.9 ± 0.44	1.0
5–cJun	67	74	2.8	-12	29 ± 19	1.1 ± 0.25	-10 ± 0.39	-13 ± 0.80	-2.6 ± 1.4	1.1
6–cJun	66	56	2.4	-12	19 ± 6.6	0.57 ± 0.01	-10 ± 0.20	-16 ± 1.0	-6.0 ± 1.1	1.0
7–cJun	71	73	3.0	-11	8.6 ± 1.8	0.99 ± 0.01	-11 ± 0.12	-14 ± 1.1	-2.8 ± 1.2	1.1
8–cJun	71	75	0.54	-12	9.9 ± 4.0	0.77 ± 0.14	-11 ± 0.23	-12 ± 1.5	-1.5 ± 1.6	1.0
9–cJun	68	77	1.6	-12	11 ± 5.5	0.60 ± 0.11	-11 ± 0.30	-14 ± 1.5	-2.9 ± 1.7	1.0
10–cJun	NF	34	NF	NF	ND	ND	ND	ND	ND	0.82
11–cJun	NF	14	NF	NF	ND	ND	ND	ND	ND	0.61
12–cJun	NF	25	NF	NF	ND	ND	ND	ND	ND	0.78
13–cJun	NF	27	NF	NF	ND	ND	ND	ND	ND	0.79
14–cJun	44	44	1100	-8.0	2200 ± 450	0.75 ± 0.29	-7.6 ± 0.12	-13 ± 1.0	-5.7 ± 1.0	0.97
15–cJun	48	62	1100	-8.0	1000 ± 560	0.48 ± 0.15	-8.0 ± 0.32	-13 ± 0.85	-5.2 ± 1.1	1.0
16–cJun	38	49	6000	-7.0	ND	ND	ND	ND	ND	1.0
17–cJun	19	44	100,000	-5.4	ND	ND	ND	ND	ND	0.85
18–cJun	60	69	18	-10	56 ± 14	0.71 ± 0.04	-9.7 ± 0.15	-14 ± 0.35	-4.0 ± 0.36	1.0
19–cJun	63	60	8.3	-11	110 ± 50	0.62 ± 0.12	-9.4 ± 0.28	-10 ± 0.54	-1.1 ± 0.53	0.98
20–cJun	47	62	2000	-7.7	2000 ± 420	0.67 ± 0.08	-7.6 ± 0.12	-4.9 ± 0.37	2.8 ± 0.40	0.99
LIN**20**–cJun	NF	39	NF	NF	55,000 ± 9700	1.4 ± 0.07	-5.7 ± 0.10	-6.0 ± 0.38	-0.25 ± 0.38	0.81
LIN**20**-TAT–cJun	42	59	3500	-7.3	ND	ND	ND	ND	ND	0.91
Pal-**20**–cJun	NF	28	NF	NF	ND	ND	ND	ND	ND	0.82
20-TAT–cJun	53	70	1100	-8.0	5600 ± 380	1.1 ± 0.01	-7.1 ± 0.04	-5.3 ± 0.05	1.8 ± 0.06	0.95
20-NLS-TAT–cJun	43	39	3100	-7.4	7600 ± 460	0.59 ± 0.00	-6.9 ± 0.04	-5.5 ± 0.06	1.3 ± 0.07	0.79
NLS-TAT–cJun	NF	11	NF	NF	NF	NF	NF	NF	NF	0.33
18HC–cJun	48	44	69	-9.6	320 ± 50	0.70 ± 0.01	-8.7 ± 0.09	-6.6 ± 0.40	2.1 ± 0.50	0.89
20HC–cJun	33	35	3300	-7.4	15,000 ± 2400	1.1 ± 0.23	-6.5 ± 0.09	-4.5 ± 0.40	2.0 ± 0.41	0.81

3.2.2.1 Single helix constraints are tolerated in uncapped peptides (1-5)

Removing helix capping motifs generated "full length" FosW derivatives **1**-**5**, featuring 10 ***a***/***d*** residues involved in core interactions and two out of a possible six electrostatic interhelical interactions between eight ***e*** and ***g*** residues (based on the FosW–cJun crystal structure). Scanning single lactam constraints through the FosW sequence revealed that four of the five positions examined improved both peptide–cJun helicity and binding affinity relative to the linear parental sequence. Constraining all heptads of FosW except Heptad 3 generated peptides that formed heterodimeric CCs with cJun (helical CD signals in Figure 3.3, and 222:208 ratios ≈1 in Table 3.1) that were more helical (+7 – +37 %) than FosW–cJun whilst being of similar or slightly higher thermal stability (T_m) (Table 3.1). Constraints induced greater helicity in the weakly helical N-terminal two heptads of FosW (peptides **5** and **4**) than in the intrinsically helical C-terminus (peptide **1**) (Mason *et al.*, 2007a), with helicity increasing as the constraint was moved towards the N-terminus.

Improved helicity was concomitant with binding affinities by ITC that were similar to FosW–cJun despite removal of helix-capping motifs, and peptides displayed similar binding affinities with constraint of N- or C-terminal heptads (Table 3.1). Thus, neither terminus was specifically identified as dispensable and a target for truncation. Instead, the poor helicity and affinity of the Heptad 3 constraint indicated a context dependence to constraint tolerance, a phenomenon that has been suggested previously (Rao *et al.*, 2013). However, the extreme C-terminal constraint of **1** resulted in the most favourable (most positive) entropic (TΔS) term (with the possible exception of **5**), being similar to that of FosW, despite **1** displaying the smallest increase in helicity relative to peptides **2**, **4** and **5** (Table 3.1). This favourable

entropic term suggested that constraints could entropically pre-organise peptides into helices for binding as hypothesized. Consequently, the constraint of **1** was retained in later designs featuring the extreme C-terminus of FosW.

Data from **1**-**5** suggested that the FosW–cJun interaction generally tolerated the introduction of helix constraints well across most of the heptads of FosW, and that constraints could be effective at inducing helicity in truncated peptides as well as in full length peptides. This might lead to retained ability to form helices and thus coiled coils with cJun for shortened peptides.

3.2.2.2 Multiple constraints improve binding affinity and helicity (6-9)

In the next cohort, doubly constrained full-length peptides were generated based on the hypothesis that truncated peptides may require multiple constraints to retain ΔG. These peptides featured the successful constraint of **1** to prepare for upcoming N-terminal truncation, and a second ***f–c*** (**6**-**8**) or ***b–f*** (**9**) constraint scanned across the molecule. Introducing two constraints generated peptides which formed heterodimeric CCs (Figure 3.3 and 3.4) with increased helicity over singly constrained peptides (**7**-**9**, with the exception of **5**), whilst again maintaining FosW-like affinities as measured by ITC (Table 3.1). Scanning one constraint progressively towards the N-terminus of the molecule slightly increased heterotypic helicity, in agreement with data from **1**-**5**. The greater helicity of **9** relative to **6**-**8** suggested the benefit of constraining extreme peptide termini, as noted for **1**-**5**. The success of this may reflect reduction of the detrimental effects of helix 'fraying' on peptide helicity (Rohl *et al.*, 1992) at constrained termini relative to linear peptides.

Data from **1**-**5** and **6**-**9** suggested that the FosW–cJun interaction is tolerant to multiple constraints across FosW, and that multiple constraints could potentially retain binding affinity in truncated peptides through helicity induction.

3.2.2.3 Constraints retain binding affinity upon peptide truncation (10-20)

Data from the above peptides were used to place two or three constraints near helix termini in N- and/or C-terminally truncated peptides.

3.2.2.3.1 N-terminal truncation and constraint (10-11)

The third peptide cohort tested whether N-terminal truncations would leave the highly helical C-terminus of FosW to generate successful cJun-binders (Mason *et al.*, 2007a), and inform as to whether appreciable helicity alone is sufficient to allow binding. However, NΔ12 truncations (relative to FosW) with the C-terminal constraint of **1**, and previously successful Heptad 2–3 (to generate **10**) and Heptad 3–4 (generating triply constrained **11**) ***f*–*c*** constraints, abolished heterodimerisation and high helicity (222:208 ratio <≈0.9, thermal denaturation was non-cooperative, and CD traces did not exceed the average, Figure 3.3 and 3.4), despite **10** heteromeric sample helicity being similar to that of FosW–cJun (Table 3.1). Peptide **10** therefore suggested helicity alone may not be sufficient for cJun binding, an observation that matches previous cFos targeting (Rao *et al.*, 2013). The constraints of **10** and **11** could not compensate for the loss of enthalpy from removal of three core ***a***/***d*** interactions and the ***g***1–***e'***2 interaction of FosW, suggesting that retention of more of the N-terminus than in these 22mers was necessary for binding.

3.2.2.3.2 The mid heptads, and C-terminal truncation and constraint (12-15)

It was hypothesized that truncation of the helical C-terminus of FosW but retention of more of the N-terminus could be successful if high molecule helicity was achieved by N-terminal constraint. Doubly- and triply-constrained NΔ5CΔ10 (relative to FosW) peptides **12** and **13** retained seven residues more of the N-terminus of FosW relative to the previous cohort, and trialled seven residues less of the C-terminus. However, neither peptide dimerised with cJun as evidenced by CD (Figure 3.3 and 3.4), despite comparable heteromeric helicity of **13** to that of FosW–cJun, again indicating that appreciable helicity alone cannot guarantee binding.

Instead, maintaining the full N-terminus (minus capping motifs) to retain eight ***a***/***d*** and both ***g***–***e'*** interactions of FosW was successful in identifying two C-terminally truncated 25mers that retained binding free energy. A doubly-constrained NΔ2CΔ10 peptide (**14**) successfully adopted heterotypic CCs of slightly higher helicity (+7 %) than that of FosW–cJun, and displayed a favourable increase in helicity (+5 %) and thermal stability with heterodimerisation relative to self-interaction (Appendix Table 3.2 and Figure 3.4). Despite losing 32 % of the molecule and a highly helical C-terminal region, and the enthalpy of two hydrophobic core ***a***/***d*** interactions, this peptide displayed a favourable retention of ΔG (77 %) compared to FosW–cJun, and a favourable increase in binding enthalpy (-2.8 kcal/mol) (Table 3.1). Compared to the previous cohort, maintaining the full N-terminus – one extra core enthalpic ***a***–***a'*** interaction – appeared to make the difference between binding and non-binding.

Adding a third constraint (**15**) further increased helicity (+18 %, now 1.7-fold higher than that of FosW–cJun), thermal stability (+4 °C) and improvement of binding

ΔG by -0.5 kcal/mol, to now retain 81 % of the FosW ΔG whilst being only 66 % of the size (Table 3.1). Thus, removing the highly helical C-terminus could be compensated for by retention of enthalpy at the N-terminus and the helicity inducing effects of constraints. This demonstrated that constraints could enable retention of binding free energy when placed appropriately and combined with suitable truncation.

3.2.2.3.3 Towards N-terminal truncation and constraint (16-20)

Next, alternative N-terminal truncation to that previously attempted was trialled, aiming to avoid loss of binding by inclusion of an extra four residues of the C-terminus. Omission of the N-terminal constraint of the successful N-terminal 25mers tested retention of sufficient helicity for binding (**16**) in the following truncation (**17**). This second peptide featured both N- and C-terminal truncation (now NΔ9CΔ6 compared to FosW) to result in a 22mer, and featured internal rather than terminal constraints to allow further truncation if successful.

With the full N-terminus (**16**), heterodimerisation was retained and helicity of the heterodimer slightly exceeded FosW–cJun; however, retention of approximately 71 % of FosW–cJun ΔG (based on CD data, Table 3.1) confirmed the necessity for the planned C-terminal extension in the next truncated peptide (**17**). Despite slightly higher helicity than FosW–cJun, **17** heterodimerisation was negligible, with a T_m of 19 °C, similar to that of native cFos–cJun. Once again, this demonstrated that truncation of three ***a***/***d*** interactions and the ***g***1–***e'***2 interaction of FosW required more extensive compensation by helicity induction than was achieved, and the benefits of constraint of molecule termini.

Based on the above data, the C-terminus of **17** was extended to full length (minus capping motifs) and combined with a more conservative N-terminal truncation at first (NΔ5) to rule out the possibility of an interaction "hot spot" at the extreme N-terminus (**18**), before stepwise internal constraint (**19**) and further N-terminal truncation to achieve a third 25mer (**20**, NΔ9CΔ3 compared to FosW). Conservative NΔ5 truncation (removal of one ***a*** residue) was successful in identifying 29mer peptides **18** and **19** forming CCs with cJun of improved helicity (+32 % and +23 %) relative to FosW–cJun. Furthermore, these peptides displayed similar thermal stability, and similar binding affinity by ITC to FosW despite being shorter (Table 3.1), including in the case of **19** a reasonably favourable entropic term. This data dismissed an N-terminal interaction "hot spot", and allowed truncation to 25mer peptide **20**.

3.2.2.3.4 Peptide 20 binds cJun effectively via entropic pre-organisation

Peptide **20**, an NΔ9CΔ3 peptide (featuring eight ***a***/***d*** interactions and the ***g***3–***e'***4 interaction of FosW), heterodimerised with cJun to create complexes 1.7-fold more helical than FosW–cJun through combination of the high native helicity of the C-terminus with N-terminal helix constraint positions effective at improving helicity. Despite truncation by one-third, **20**–cJun displayed a T_m 10 °C above physiological temperature, and maintained approx. 77 % of the free energy of FosW–cJun interaction despite being only 66 % of the length of FosW (Table 3.1). It is noted that this ΔG retention is approximate due to the relatively flat line shape of the ITC titrations at the peptide concentrations assayed (low titration c value, see also 2.9.7). Approximately 5.7 kcal/mol of binding enthalpy was lost in truncation of 12 residues including helix-capping motifs and two ***a***/***d*** and one ***g***–***e'*** interhelical interactions. However, this enthalpic loss was largely compensated for by a considerable gain to

the entropic binding contribution (Δ[TΔS] approx. +3.2 ± 0.5 kcal/mol), enabling this retention of free energy. Thus, **20** displayed for the first time the expected entropic benefit of constraining (Rao *et al.*, 2013), which in previous peptides had been precluded by retention instead of the favourable binding enthalpy of FosW. This peptide demonstrated that the N-terminus of FosW is not absolutely required for cJun interaction, and that N-terminal truncation could be successful with correct constraint placement, a theme common amongst the peptides in this study, and previously noted in a similar design strategy (Rao *et al.*, 2013).

Compared to its unconstrained counterpart LIN20 with cJun, **20**–cJun has far higher helicity (62 % vs. 38 %); though helicity of LIN20–cJun is comparable to that of FosW–cJun (37 %), LIN20 in contrast to **20** and FosW displayed poor heterodimerisation according to CD experiments, with a T_m approximately 22 °C lower than **20**. ITC analysis indicated that LIN20 bound cJun more weakly than cFos (Table 3.1). Further, LIN20 lost approximately 4.4 kcal/mol of binding enthalpy (again noting that the flat line shape of the LIN20 titration), through truncation of 12 residues including helix-capping motifs and two ***a***/***d*** and one ***g***–***e'*** interhelical interactions. This is comparable to the enthalpy loss of **20** (≈5.7 kcal/mol) with the same truncation; however, without constraints, LIN20 was not able to compensate for this enthalpy loss through an improved entropic term to binding, and so as a result lost approx. 43 % of binding free energy. Peptide **20** on the other hand, with its lactam constraints, retained slightly more of the binding free energy of FosW relative to LIN20.

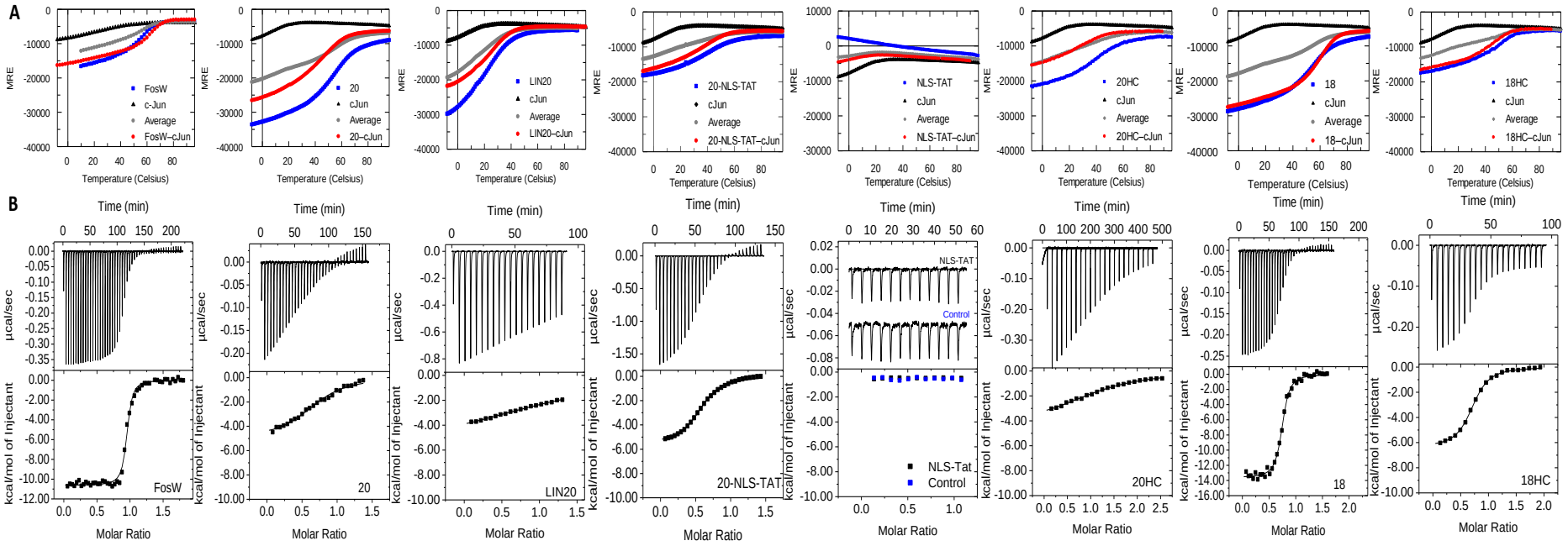


Figure 3.6: Peptides 18 and 20 retain binding free energy of FosW despite truncation and are tolerant to cell penetrating moiety attachment or hydrocarbon constraint replacement. Thermal denaturation profiles (**A**) and isothermal titration calorimetry (**B**) of FosW, **20**, LIN20, **20**-NLS-TAT, NLS-TAT, **20**HC, **18** and **18**HC in interaction with cJun. **A** Thermal melt data is reported as change in mean residue ellipticity (MRE; units deg cm^2/dmol), as a function of circular dichroism ellipticity at 222 nm with temperature. Blue squares: peptide alone, red circles: equimolar mix of peptide with cJun, grey diamonds: average of cJun and peptide alone, black triangles: cJun alone. **B** Raw ITC isotherms (top panels) and fitted data (bottom panels) (both baseline corrected). Injection enthalpies for the buffer into buffer control (blue text label) on the NLS-TAT plot are translated by -0.05 µcal/sec for clarity relative to those for cJun into NLS-TAT.

In summary, three truncated 25mer peptides demonstrated that the loss of enthalpy and binding free energy associated with considerable truncation from either terminus of FosW can successfully be limited by the beneficial effects of appropriately placed helix constraints. In particular, peptide **20** displayed the hypothesized entropic benefit of constraining that resulted in retention of the majority of the FosW binding free energy with truncation by one-third.

3.2.3 Helicity, entropy and binding affinity of constrained/truncated peptides

Helicity is expected to be an important property of peptides able to supercoil with cJun (see 3.3.6 for discussion). Previously derived cJun antagonists and native cJun binders are predicted by the Agadir algorithm to have low monomeric helicities (Appendix Table 3.3) ((Lacroix *et al.*, 1998) and references therein). However, these peptides favourably form CCs with cJun that are more helical than the sum of monomeric peptide helicities, such that heterodimer helicity cannot be accurately predicted from this property (Worrall and Mason, 2011). There is a positive but poor correlation (R^2 = 0.24) between peptide K_d values by ITC and heterodimer helicity (Appendix Figure 3.1A). Interacting peptides had heteromeric helicity ≥44 % with a mean helicity (56 %) much higher than FosW–cJun (37 %), and were more helical than expected for locking residues between those forming the helix constraint into a helical conformation (Appendix Table 3.4), whilst non-interacting peptides had a mean helicity of 26 % and were less helical than expected. As helicity of peptide–cJun complexes increased, peptide affinity for cJun increased, though to a lesser extent; for example, comparing **15** and **14**, an extra 19 % helicity from the extra constraint of **15** translated to a 6 % increase in binding free energy retention.

The relationship between helicity and entropic/enthalpic contributions to binding is even more complex. Favourable enthalpy correlates positively but poorly (R^2 = 0.22) with helicity, whilst counter to what is expected, favourable entropy negatively correlates with helicity (though poorly, R^2 = 0.13) (Appendix Figure 3.1B and C). Peptide comparisons highlight the complexity of these relationships. For example, C-terminally constrained **1** had the most favourable entropy of full-length peptides **1**-**5**, despite displaying the lowest increase in heterodimer helicity vs. FosW. The second most favourable entropy in peptides **1**-**5** was that of N-terminally constrained **5** with cJun, which conversely displayed the highest increase in heterodimer helicity vs. FosW. Finally, **9**, which featured combination of the constraints of **1** and **5**, resulted in the highest heterodimer helicity in the whole dataset, but had an entropy no more favourable than **1** or **5**.

3.2.4 Exploring cellular uptake of 20, and testing 20 activity *in cellulo*

Successfully truncated 25mer peptide **20** represented the best compromise between truncation and substantial retained affinity for cJun, and so was chosen for further exploration of potential for future AP-1-targeted therapy. Investigation and optimisation of the cellular uptake of **20** was attractive to probe whether **20** displayed desirable *in cellulo* AP-1 inhibitory activity in addition to its *in vitro* cJun binding affinity. In cell assays, unmodified **20** was assayed based on the hypothesis that lactam constraint and truncation might confer membrane penetration potential, as has been reported for other constraint chemistries (Walensky *et al.*, 2004). It was expected that appendage of cell membrane- and nuclear-uptake moieties to **20** could facilitate cell uptake where unmodified **20** uptake was negligible.

Before cell assays, it was important to confirm that these moieties did not disrupt binding of **20** to cJun. Moieties tested were an N-terminal palmitic acid (Pal-**20**), a lipid attached to proteins *in vivo* for anchoring to the lipid bilayer; a C-terminal Tat peptide (**20**-Tat), a cationic Cell Penetrating Peptide (CPP) that is the minimal sequence for HIV-1 internalisation into T-cells (Debaisieux *et al.*, 2012); and novel combination of the Tat peptide with a C-terminal monopartite SV40 large T-antigen nuclear localisation signal (NLS) (**20**-NLS-Tat) (Lanford *et al.*, 1986).

The CD trace of Pal-**20** with cJun exceeding that of the average, indicating possible interaction, though the 222:208 ratio was <≈0.9 (Figure 3.3). However, the heteromeric melt lacked the usual sigmoidal shape suggestive of cooperative unfolding, making interaction difficult to ascertain (Figure 3.4), and perhaps reflecting contribution of palmitate chain interactions to the CD signal. **20**-Tat and **20**-NLS-Tat desirably displayed binding similar to that of **20** alone by CD and ITC (Figure 3.6 and Table 3.1), indicating that NLS-Tat appendage did not disrupt binding of **20**. This was corroborated by analysis of NLS-Tat, which did not display detectable binding to cJun, with a melt trace that virtually overlaid the average (Figure 3.3 and 3.4), and ITC heat spikes similar to buffer titration into buffer (Figure 3.5 and 3.6). Thus, Tat and NLS-Tat appendages allowed testing of cellular delivery of **20**.

3.2.4.1 20-NLS-Tat cancer cell uptake and nuclear localisation

To investigate cellular uptake of **20** and Tat/NLS-Tat derivatives, these peptides and control Tat peptide were synthesized with fluorescein isothiocyanate (FITC) probes attached via flexible aminohexanoic acid linkers. FITC-peptides were synthesized and purified by Dr. Timothy A. Hill, and all cell assays were performed by

Mr Samuel Perry, both at the Institute for Molecular Biosciences, University of Queensland, Australia. Fluorescence-activated cell sorting (FACS) was used to quantify intracellular fluorescence in breast cancer MCF-7 cells arising from internalised peptides FITC-**20**, FITC-**20**-Tat and FITC-**20**-NLS-Tat relative to fluorescein-Tat (FLU-Tat) control (Figure 3.7A). AP-1 is overexpressed in many human breast cancer cells including MCF-7 (Chen *et al.*, 1996), which are dependent on AP-1 activity for proliferation, and thus AP-1 is an important target for future cancer therapy (Liu *et al.*, 2002). Relative fluorescence of FITC-peptides was recorded for the entire population of treated cells, of which ≈98 % internalised peptides (see Appendix Figure 3.2 for representative example). Without additional cell penetrating moieties, there was negligible internalisation of FITC-**20** after incubating for one or six hours with cells. However, C-terminal addition of Tat (FITC-**20**-Tat) promoted attractive internalisation relative to FLU-Tat alone, particularly after six hours, and internalisation was confirmed by live cell confocal microscopy of the treated cell population (Figure 3.7B). This evidenced cytoplasmic and perinuclear punctate localisation with a low level of nuclear localisation for FITC-**20**-Tat that was somewhat similar to that of the control FLU-Tat peptide. Further addition of a nuclear localisation signal (FITC-**20**-NLS-Tat) promoted more efficient cellular uptake and nuclear localisation; more FITC-**20**-NLS-Tat was internalised at one hour and at six hours compared to FITC-**20**-Tat, resulting in an intense, diffuse cytoplasmic and nuclear localisation (Figure 3.7B). Thus, C-terminal appendage of Tat, and in particular NLS-Tat, led to efficient cellular uptake and nuclear localisation of **20**.

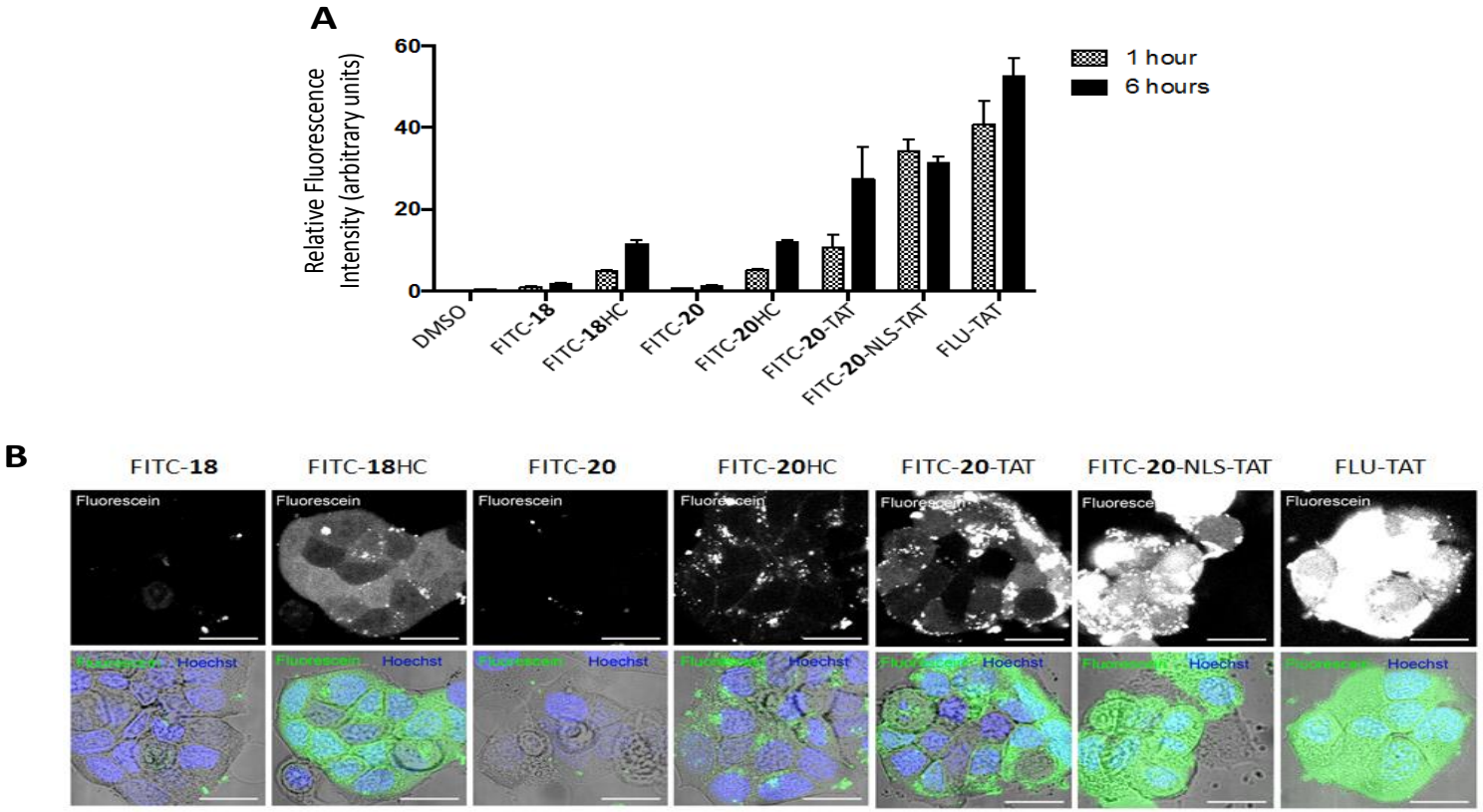


Figure 3.7: Cellular uptake and localization of 20 conjugates and hydrocarbon constraint derivative peptides in MCF-7 breast cancer cells. A Treatment of MCF-7 cells with 10 µM FITC-peptides for one hour and six hours, and quantitation of peptide uptake from the entire treated cell population by flow cytometry. Fluorescence intensities are means (+ SEM) of three independent treatment experiments. **B** Treatment of MCF-7 cells with 10 µM FITC-peptides for six hours and visualisation of peptide uptake by live cell confocal microscopy. Cell nuclei were counterstained with Hoechst 33342 dye. Scale bar (white) represents 20 µm.

3.2.4.2 NLS-Tat inhibits *MMP9* and *Cyclin D1* expression *in cellulo*

Effective delivery of **20** to the nucleus of MCF-7 cells encouraged investigation as to whether this peptide could inhibit AP-1 in these cells. The ability of **20**-conjugates and control peptides (without FITC probes) to inhibit AP-1-mediated *MMP9* (Smith *et al.*, 1999) and *Cyclin D1* (Liu *et al.*, 2004) target gene transcription in treated MCF-7 cells was therefore evaluated using quantitative real-time PCR of reverse transcribed mRNA (Figure 3.8A and B). MCF-7 cells were pre-incubated with 10 μM of **20**, **20**-Tat, **20**-NLS-Tat or control NLS-Tat peptides, followed by addition of phorbol 12-myristate 13-acetate (PMA, also known as TPA) to activate AP-1 (Angel *et al.*, 1987). Preliminary qRT-PCR data is displayed in Figure 3.8A and B, and it is noted that in the future increased numbers of biological replicates will allow better evaluation of the significance of suggested activities of peptides. Pre-incubation with **20**-NLS-Tat reduced PMA-induced *MMP9* (significantly by Student's T-test analysis) and possibly *Cyclin D1* mRNA expression, whereas **20** alone showed no effect, and **20**-Tat had effects similar to **20**-NLS-Tat. However, undesirably the control peptide NLS-Tat displayed significant reduction of both *MMP9* and *Cyclin D1* expression, indicating that the suggested activity of **20**-NLS-Tat could be attributable to the NLS-Tat appendage rather than AP-1 inhibition by **20**. This indicated that NLS-Tat was inappropriate for delivery of **20** into breast cancer cells to test AP-1 inhibitory activity.

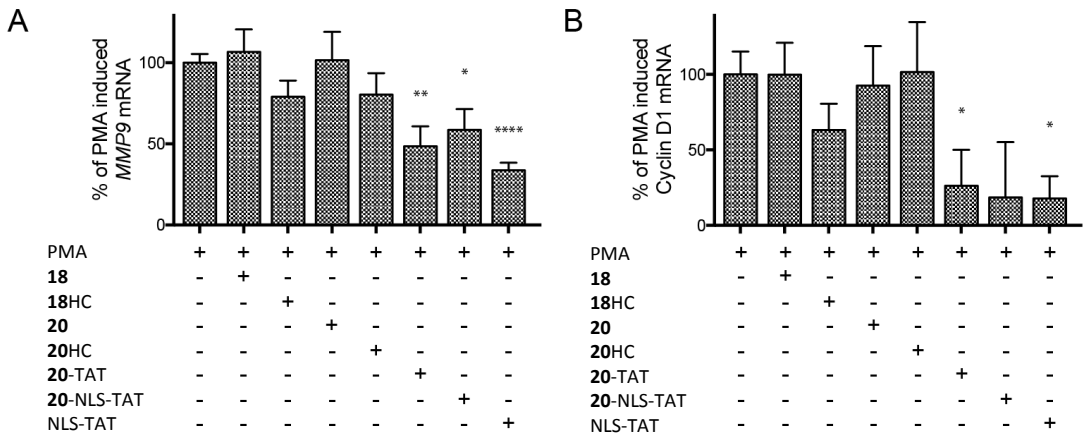


Figure 3.8: 18HC displays potential inhibition of AP-1-mediated gene expression in MCF-7 breast cancer cells. qRT-PCR analysis of *MMP9* (**A**) and *Cyclin D1* (**B**) gene expression in treated MCF-7 cells. Peptides (without FITC probes) were incubated with cells at 10 μM for 14 hours before addition of PMA for six hours. Gene expression is expressed as % of PMA-induced mRNA levels relative to vehicle (DMSO) control. Data shown are means (+ SEM) of three independent experiments. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$ (Student's T-test).

3.2.4.3 Replacement of lactam for hydrocarbon constraints in 20

Following the failure of NLS-Tat appendage for testing *in cellulo* AP-1 inhibitory activity of **20**, an alternative strategy was investigated. This involved substitution of lactam *i→i+4* helix constraints for hydrocarbon 'staples' generated by olefin metathesis (Schafmeister *et al.*, 2000), an effective constraint modality for development of PPI-modulatory peptides with drug-like properties (Robertson and Jamieson, 2015). Lactam helix constraints had initially been more attractive for probing positions in FosW tolerant to helix constraint and stabilising a large number of truncated peptide derivatives due to the low cost and ease of cyclisation of Lys and Asp side chains compared to hydrocarbon synthesis. Further, lactam helix constraints were expected to induce higher peptide helicity than hydrocarbon staples based on a recent comparative study (de Araujo *et al.*, 2014). However, in some cases cell penetration by hydrocarbon stapled peptides has been reported

(Bird *et al.*, 2014), attributed to interaction of the hydrocarbon staple with cell membrane components/internalisation mechanisms to facilitate transfer of the constrained peptide into the cytosol (Chu *et al.*, 2015). Thus, hydrocarbon staples were placed at positions known to be effective at inducing helicity from lactam constraint probing, to investigate whether stapled FosW derivatives would be cell permeable. Hydrocarbon stapled peptides were synthesized and purified by Dr. Timothy A. Hill, and cell assays performed by Mr Samuel Perry, both of the Institute for Molecular Biosciences, University of Queensland, Australia.

Replacement of lactam constraints in **20** generated peptide **20**HC (Figure 3.2). The **20**HC heteromeric mixture overlaid the average in CD scans and melts such that binding was not supported, despite comparable helicity to FosW–cJun (Figure 3.3, 3.4 and 3.6). However, binding was confirmed by ITC analysis, with **20**HC retaining 65 % of the FosW binding free energy rather than 77 % for **20** (Figure 3.5, 3.6 and Table 3.1). **20**HC displayed similar enthalpic loss and entropic gain to **20**, though it is noted that the 30 µM solubility of **20**HC in KPP buffer for ITC likely limited the accuracy of thermodynamic parameter measurement. Finally, **20**HC had a similar affinity to LIN20 by ITC.

FACS and fluorescence microscopy (Figure 3.7) were performed to evaluate cell uptake of FITC-labelled **20**HC. FITC-**20**HC exhibited low cellular uptake, approximately 20 % of that of the FLU-Tat control after 6 hours, much less than had been achieved with Tat and NLS-Tat appendages but more than FITC-**20** alone. However, microscopy evidenced that uptake was cytosolic and punctate without detectable nuclear uptake, and extracellular aggregates were also observed (Figure

3.7B). Finally, *MMP9* and *Cyclin D1* expression levels in MCF-7 cells treated with **20**HC were measured to test for AP-1 inhibition (Figure 3.8); however, in this preliminary data, reductions in levels of both genes were negligible. Thus, hydrocarbon stapled **20**HC was able to enter cells, but to a low extent, and once in cells nuclear uptake was not demonstrated and nor was inhibition of AP-1-mediated gene expression.

3.2.5 Hydrocarbon stapled 18HC enters cells and potentially inhibits AP-1

In addition to a hydrocarbon version of **20**, hydrocarbon staples were also substituted for lactam constraints in **18**; this peptide had previously shown much higher affinity for cJun than **20**, though **20** had originally been preferred as it is four amino acids shorter. The resultant **18**HC was evaluated for improved binding to cJun by CD and ITC (Figure 3.3 – 3.6 and Table 3.1), cellular uptake (Figure 3.7) and AP-1 target gene expression inhibition (Figure 3.8) relative to **20**HC, to determine whether it represented more attractive properties for these goals.

18HC heterodimerised with cJun (222:208 ≈0.9) to create complexes of comparable helicity to FosW–cJun, though lesser than **18**–cJun, supporting greater helicity induction by lactam constraints than hydrocarbon staples (de Araujo *et al.*, 2014). **18**HC–cJun complexes were less thermally stable (T_m 15 °C lower) than FosW, but were more thermally stable than **20**HC–cJun by the same margin, and ITC determined **18**HC to bind cJun with a free energy 88 % of that of FosW, and just below that of **18**, despite truncation by 22 % relative to FosW. Interestingly, **18**HC displayed a favourable entropic term to binding ($\Delta T\Delta S$ = 2.6 ± 0.7 kcal/mol relative to FosW) where **18** had not. Thus, **18**HC retains almost identical binding free energy for its

truncation vs. FosW as **20** does, whilst being a nanomolar affinity cJun antagonist only slightly longer than micromolar affinity binder **20**.

FACS and fluorescence microscopy (Figure 3.7) demonstrated similar uptake quantities of FITC-**18**HC to those of FITC-**20**HC after one and six hours, but in contrast to FITC-**20**HC, FITC-**18**HC displayed attractive diffuse and intense cytosolic and nuclear staining similar to that of FITC-**20**-NLS-Tat. FITC-**18** was also assayed, but like FITC-**20** was not taken up by cells. Thus, FITC-**18**HC demonstrated desirable uptake and subcellular localisation properties. Preliminary data further suggested that this peptide reduced *MMP9* and *Cyclin D1* expression levels in MCF-7 cells (Figure 3.8). However, at this point, these reductions were not statistically significant (Student's T-test), and further investigation involving repetition of this assay with more biological replicates is needed to corroborate this suggestion.

In summary, through a series of peptides investigating different truncations and helix constraint modalities, FosW has been truncated by 22 % with retention of 88 % of binding free energy through the use of stabilising hydrocarbon staple helix constraints to generate peptide **18**HC. This peptide has attractive nanomolar affinity for cJun *in vitro*, and desirable cell permeation and subcellular localisation properties. Finally, it is suggested that this peptide may inhibit AP-1 activity in breast cancer cells.

3.2.6 Comparison of 18HC against other AP-1 inhibitory agents

Compared to other helical LZ CC domain antagonists 4hFosW and $FosW_{CANDI}$ (Crooks *et al.*, 2011, Mason *et al.*, 2007b), **18**HC forms CCs with cJun of a lower helicity but comparable thermal stability to both peptides. The affinity of **18**HC for cJun by ITC is almost identical to that of 4hFosW whilst being 12 % shorter, and

comparable to $FosW_{CANDI}$ (extrapolating from T_m values) whilst being 22 % shorter. The helicity, thermal stability and affinity of **18**HC is much higher than native cFos–cJun (Worrall and Mason, 2011), such that **18**HC would be expected to outcompete cFos for cJun *in cellulo* to effectively antagonise cFos–cJun AP-1 formation.

The shortest LZ CC domain antagonist generated to date is a 22mer constrained and truncated peptide against cFos, derived from $JunW_{CANDI}$ (Rao *et al.*, 2013). **18**HC retained virtually identical % parental ΔG to this 22mer, but in a molecule 19 % longer vs. FosW. However, the binding affinity of **18**HC for cJun (320 ± 50 nM by ITC) is much higher than the $JunW_{CANDI}$ derivative for cFos (7.3 ± 0.64 μM) such that lower concentrations are required to inhibit AP-1 formation efficiently, and **18**HC has additionally demonstrated cellular uptake and potential *in cellulo* activity.

Peptide **18**HC also compares well with other AP-1 antagonists derived by alternative strategies to that described in this Chapter, such as relatively short existing peptidic AP-1 antagonists (see 1.3.2). For example, the affinity of **18**HC is slightly lower than that of computationally designed Jun-d1 (6 nM as determined by solution fluorescence resonance energy transfer (FRET)), and substantially lower than peptide "A-Fos" whose acidic extension to the LZ CC domain sequesters the cJun basic DNA-binding domain (0.03 nM as estimated by CD thermal denaturation), whilst being more attractive than peptide-DNA conjugate C2ds (362 nM as determined by fluorescence anisotropy), assuming that affinities measured by different biophysical techniques can be compared equally. However, all three of these molecules are also longer and thus arguably less therapeutically attractive than **18**HC; at 29 residues in length **18**HC is shorter than Jun-d1 by 13 residues, shorter

than C2ds by 7 residues and lacking 18 bp of conjugated dsDNA, and shorter than A-Fos by 34 residues. Thus, the affinity of **18**HC is appreciable given its smaller size relative to these molecules, and its smaller size is expected to improve therapeutic viability. **18**HC is also much shorter and more attractive for further development than long transactivation domain deletion mutants like 210mer TAM67.

Finally, **18**HC compares well with small molecules under development against AP-1. These include DNA-binding inhibitors T-5224 and momordin I (see 1.3.1), which have displayed an *in cellulo* IC_{50} of ≈10 μM and an *in vitro* IC_{50} of 30 μM respectively (Aikawa *et al.*, 2008, Park *et al.*, 2000). Though the suggested *in cellulo* activity of **18**HC awaits confirmation, the attractive *in vitro* K_d of this peptide is promising for translating to a desirable *in cellulo* IC_{50} relative to these small molecules.

3.3 Discussion

3.3.1 Helix constraint and truncation of FosW

FosW has high affinity for cJun, but is too long to be therapeutically viable, such that truncation to improve drug-likeness was desirable. However, it was hypothesized that truncation could abolish binding, such that helix constraint to reinforce helical structure could generate short peptides retaining sufficient affinity.

Helix constraints encourage peptides into α-helical rather than random coil conformations, an event disfavoured by a loss of conformational entropy (S_{conf}) as residue conformation ranges are limited relative to the random coil state (Zimm and Bragg, 1959). In longer peptides of residues with sufficiently high α-helical propensities (O'Neil and Degrado, 1990), this entropic penalty can be overcome by a

more favourable enthalpic gain from backbone hydrogen bonds in the α-helix, such that the overall ΔG of helix formation is favourable. In shorter peptides, this hydrogen bonding network is reduced, such that the unfavourable S_{conf} predominates and the peptide adopts random coil conformations. Helix constraints are covalent cyclisations of amino acids that lock the backbone of intervening residues in an α-helical conformation. In doing so, a constrained "nucleus" is created from which the rest of the peptide can adopt an α-helix with a reduced (or 'prepaid') S_{conf} entropic penalty (Zimm and Bragg, 1959, Harrison *et al.*, 2010, Rao *et al.*, 2013).

cJun binding is dependent on supercoiling event of helical partners (Crick, 1953), such that constraints that increase helix adoption are expected to increase the affinity of peptides for supercoiling (O'Neil and Degrado, 1990). Upon peptide truncation and loss of helical structure, constraints that entropically pre-organise peptides into helices relative to unconstrained peptides may therefore retain binding affinity of longer peptides despite loss of enthalpic interactions (Rao *et al.*, 2013).

Helix constraints have successfully achieved entropic stabilisation in various helical systems including that of AP-1 (Harrison *et al.*, 2010, Rao *et al.*, 2013), where *i→i+4* lactam constraints enabled truncation of cFos antagonist $JunW_{CANDI}$ by 40 % whilst increasing helicity by 10 % and retaining 87 % of binding free energy (Rao *et al.*, 2013). Lys-to-Asp lactam constraints at *i→i+4* positions favour α-helix adoption by constraining residues one helical turn apart (Pauling *et al.*, 1951), a distance spanned well by Lys–Asp but not by alternative side chains such as Lys–Glu (Shepherd *et al.*, 2005), to encourage an intervening α-helix backbone geometry that is near-

ideal (Hoang *et al.*, 2016). This success encouraged application of ***b–f*** or ***f–c*** *i→i+4* lactam constraints to FosW before combining constraint with truncation.

Without helix capping motifs to participate in otherwise unfulfilled backbone hydrogen bonds at FosW termini (Doig and Baldwin, 1995, Chakrabartty *et al.*, 1993a), helix constraints conferred affinities to 32mer peptides not substantially higher than FosW as expected from entropic gain, but did improve helicity and demonstrate the previously unknown tolerance of FosW to constraint insertion. Truncated peptides with nevertheless higher heterodimeric helicity than FosW could be generated by constraint and truncation from either terminus, but only with conservative truncation and constraint of peptide termini, limiting further truncation. C-terminal constraint was likely successful via reducing more extensive fraying at this terminus (Chakrabartty *et al.*, 1993a) and greater propagation of helicity towards the N-terminus from lactam constraints (Hoang *et al.*, 2016), whilst N-terminal constraint was likely successful due to the intrinsically low helicity of this region of FosW (Mason *et al.*, 2007a). Constraint success was also sequence-context dependent, perhaps due to interactions of neighbouring side chains with the constraint, which could include H-bonding with the lactam bond of the constraint.

N-terminally truncated peptide **20**, one-third shorter than FosW but retaining 77 % of FosW binding free energy through helix-constrained entropic pre-organisation, was initially the best compromise between truncation achieved and affinity retained. As expected, in LIN20 the same truncation without constraint readily attenuated binding enthalpy and therefore affinity (De Crescenzo *et al.*, 2003), demonstrating the necessity of constraint for maintaining more attractive

binding affinity. Constraints may also improve protease resistance of FosW derivatives, as demonstrated by substantial improvement of $JunW_{CANDI}$ half-life in human serum (Rao *et al.*, 2013).

Finally, the loss of enthalpy from truncation of two ***a***/***d*** interactions (and lesser enthalpy loss from one ***g***–***e'*** interaction) that limited further truncation of FosW indicates that binding enthalpy is widely distributed along the core of cJun. Lack of binding "hot spots" is a common problem for targeting of PPIs generally (Corbi-Verge and Kim, 2016), and has been concluded previously for cFos (Rao *et al.*, 2013). This could be a contributing factor to why few successful constrained helical peptide inhibitors of CCs have been reported in the literature (Edwards *et al.*, 2016). Furthermore, binding specificity would be expected to be similarly spread along cJun core and ***e***/***g*** positions, again limiting truncation of antagonists. These factors adds to the difficulty of targeting AP-1, already considerable given that AP-1 activity is localised in the nucleus and so AP-1 is less accessible to therapeutic agents than cytoplasmic or extracellular proteins (Craik *et al.*, 2013).

3.3.2 NLS-Tat appendage for cellular uptake and nuclear localisation of 20

Helix constraints can in some cases also improve cellular uptake of peptides (Bernal *et al.*, 2007, Walensky *et al.*, 2004), and so were attractive modifications for peptides targeting nuclear AP-1 activity. Cellular uptake of constrained peptides is nevertheless difficult to predict and design, despite considerable investigation (Chu *et al.*, 2015, Bird *et al.*, 2014). It has emerged that constrained peptide internalisation is highly sequence- and cell-dependent (Bird *et al.*, 2014). In addition, the focus of many of these studies on hydrocarbon staples meant that cellular uptake by lactam

constrained FosW derivatives was not guaranteed. Indeed, whilst the main aim of entropically pre-organising **20** for binding cJun was achieved with lactam constraints, internalisation of FITC-**20** into MCF-7 breast cancer cells was negligible (Figure 3.7).

The extent of binding free energy retention of **20**, however, made testing *in cellulo* activity of this peptide an attractive further aim, and so appendage of cell- and nuclear-penetrating moieties aimed to improve cellular and nuclear uptake of **20**. Whilst these additions did increase the molecule length, the active unit (**20**) remained a more attractive molecule than longer antagonists, which would likely require similar modification for delivery. The ability of the HIV-1 Tat CPP to enter cells through a variety of pathways (Debaisieux *et al.*, 2012), and presence of an NLS within its sequence, have seen its wide use for the cellular uptake of various 'cargo' molecules (Dietz and Bahr, 2004). However, the predominately punctate localisation of FITC-**20**-Tat suggested the majority of peptide was trapped within the endosomal system (Figure 3.7). This is a common and non-trivial hurdle for CPP-cargo complexes generally (Dietz and Bahr, 2004). For instance, previously, approx. 50 % of Tat incubated with HeLa cells at 7 μM became trapped within endosomes, with the remaining fraction distributed between cytoplasm and nucleus (Potocky *et al.*, 2003), though increased endosomal escape has been reported at higher concentrations (Duchardt *et al.*, 2007). Based on these studies, more cytoplasmic and nuclear FITC-**20**-Tat was expected here. In addition, besides endocytosis Tat and other Arg-rich CPPs can directly translocate across the cell membrane via interaction with anionic phospholipid head groups and formation of transient pores (Herce *et al.*, 2014), such that a proportion of FITC-**20**-Tat would be expected to enter the cytosol directly and avoid endosomal trapping. Because FITC-**20**-Tat is still much below the size limit of

the nuclear pore complex, peptide that is able to access the cytosol could then diffuse into the nucleus to reach AP-1 (Debaisieux *et al.*, 2012). Though nuclear delivery of **20** was most desirable, cytoplasmic co-/post-translational cJun sequestration is also attractive as this is likely to expedite cJun proteasomal degradation to prevent AP-1 formation. Despite these arguments, the observed inefficiency of Tat for nuclear delivery of **20** warranted further optimisation.

To improve nuclear uptake, the SV40 large T-antigen NLS PKKKRKV was appended between **20** and Tat (Lanford *et al.*, 1986). This NLS has not been used with Tat extensively for peptide delivery, though a similar NLS (RKRRK) with Tat enabled nuclear delivery of β-catenin targeting peptides (Hsieh *et al.*, 2016). Endosomal escape or direct transduction into the cytosol was apparently improved by addition of the SV40 NLS, to deliver **20** to the cytoplasm and nucleus (Figure 3.7) at 10 μM, which compared to an *in vitro* K_d of approx. 2.0 ± 0.4 μM meant **20** was likely delivered at an effective concentration for AP-1 inhibition (Bains *et al.*, 1997).

3.3.3 Undesirable inhibition of *MMP9* and *Cyclin D1* expression by NLS-Tat

AP-1 drives expression of extracellular matrix proteinase MMP9 and G1→S phase regulator Cyclin D1 in MCF-7 breast cancer cells, repression of which reduces metastasis and proliferation making these genes particularly relevant for monitoring AP-1 inhibition (Smith *et al.*, 1999, Liu *et al.*, 2004). Unfortunately, NLS-Tat alone significantly inhibited AP-1 target gene expression (Figure 3.8).

The exact mechanism for this activity, and whether NLS, Tat or NLS-Tat together are responsible, is currently unclear. Possible mechanisms include (but are not limited to) direct binding and inhibition of AP-1, or of upstream activators or

transcription coactivators, or cytotoxicity. Of these possibilities, direct inhibition of AP-1 is unlikely as NLS-Tat does not bind cJun significantly *in vitro* (Figures 3.3 – 3.6). Recently, Tat alone was shown to marginally decrease sarcoma cell viability at 10 µM (the concentration used here for qRT-PCR) (Li *et al.*, 2014), suggesting cytotoxicity as a possible cause here. However, this report contradicts the use of Tat at up to 100 µM in a wide variety of cell lines without reported toxicity (Dietz and Bahr, 2004). Tat with a shorter NLS also did not reduce viability of two normal cell lines at 100 µM (Hsieh *et al.*, 2016), and the NLS used here alone was not toxic to a HeLa derivative cell line up to 100 µM (Sibrian-Vazquez *et al.*, 2010). The effects of NLS-Tat reported here will therefore require further experimental investigation.

3.3.4 Hydrocarbon stapled 18HC enters cells to potentially inhibit AP-1

The previously reported cellular uptake of some hydrocarbon stapled peptides prompted replacement of lactam constraints for hydrocarbon constraints in FosW derivatives, with the aim of improving uptake without the need for NLS-Tat or similar appendages. The exact mechanisms and peptide properties involved for stapled peptide cell penetration are still unclear (Chu *et al.*, 2015). For instance, it has been argued that stapled peptides may bind to the interface of cell membrane lipid bilayers causing membrane stretching, thinning and transient pore formation to allow direct passive diffusion into the cytosol (Sun *et al.*, 2013). This perhaps is afforded through shielding of the hydrophilic backbone amides/carboxyls in the stapled peptide from solvation, and provision of a hydrophobic surface (the staple itself) to interact with the membrane (Chu *et al.*, 2015). It has also been suggested that stapled peptides may enter cells via energy-dependent (ATP-driven) endocytosis

mechanisms, potentially those not involving clathrin and caveolin and via interaction with negatively charged membrane-anchored proteoglycans, as determined in HeLa cells and a normal cell line (Chu *et al.*, 2015). This incomplete understanding means replacement of lactam for hydrocarbon constraints was not guaranteed to improve cellular uptake of FosW derivatives. Nevertheless, this strategy was trialled through modification of **20** (**20**HC) and also slightly longer but much higher affinity **18** (**18**HC).

Hydrocarbon stapled FITC-**20**HC unfortunately appeared to be predominantly trapped in endosomes or involved in intracellular aggregates in MCF-7 cells (Figure 3.7). Hydrocarbon constrained FITC-**18**HC, on the other hand, was much more promising, demonstrating appreciable cytoplasmic and nuclear staining despite net uptake to a similar low level as FITC-**20**HC. Furthermore, **18**HC demonstrated potential for reduction of AP-1 transcriptional activity, which will be confirmed in follow up experimentation. Though longer than originally desired at 29 residues, **18**HC is still appreciably shorter than FosW, and has much improved cJun affinity (320 ± 50 nM) relative to **20** (approx. 2.0 ± 0.4 μM) resulting from entropic pre-organisation to make **18**HC more attractive for *in vitro* binding. Moreover, **18**HC penetrated breast cancer cells to inhibit AP-1 without the undesirable effects of NLS-Tat that was required for **20** uptake, and so was the most attractive peptide derived.

3.3.5 Hydrocarbon stapled 18HC is an attractive AP-1 antagonist

18HC compares well with previous described AP-1 antagonists. Though the truncation achieved was not as extensive as for the $JunW_{CANDI}$ derivative of Rao *et al.* (2013), **18**HC is expected to be a more valuable antagonist due to targeting cJun rather than cFos. cJun features a more typical hydrophobic core (with only one polar

a position N) than cFos, whose core contains two polar Thr and two charged Lys ***a*** residues, such that the maximal binding ΔG that can be achieved in complex with an antagonist is lower than for cJun (Acharya *et al.*, 2006). As a result of this core, cFos is thought to be unable to form homodimers (Smeal *et al.*, 1989) and so cannot initiate transcription and oncogenesis without heterodimerising with cJun (Chiu *et al.*, 1988, Schuermann *et al.*, 1989). Finally, cJun-containing complexes are some of the most potent DNA-binding and transactivating AP-1 compositions (Ryseck and Bravo, 1991, Yang-Yen *et al.*, 1990, Chiu *et al.*, 1989). Thus, cJun antagonists like **18**HC could be more valuable for cancer therapy via AP-1 inhibition than cFos antagonists.

Furthermore, **18**HC may represent a more specific antagonist than small molecules momordin I and T-5224. Both molecules target the Arg- and Lys-rich DNA-binding domains of AP-1 component proteins (Aikawa *et al.*, 2008, Park *et al.*, 2000). Whilst these domains are different to other DNA-binding motifs like zinc fingers and helix-turn-helix motifs (Landschulz *et al.*, 1988a), homology between functionally diverse bZIP TFs is very high (Fujii *et al.*, 2000). For example, within the Fos/Jun family, Fos homologue DNA-binding domains differ by only three conservative mutations over a stretch of 20 residues, as do Jun homologues. This makes specific targeting of certain AP-1 compositions such as oncogenic cFos–cJun particularly difficult. Conversely, the LZ CC domains of Jun and Fos proteins targeted here are more dissimilar across the bZIP family than DNA-binding domains, aiding specific targeting of certain homologues (Vinson *et al.*, 2002). Despite this, T-5224 shows little inhibitory activity against bZIP ATF-2 and C/EBPα proteins (Aikawa *et al.*, 2008), indicating some specificity towards the cFos–cJun it was designed against; however, T-5224 has not yet been tested on JunB-, JunD-, Fra1- or Fra2-containing AP-1.

3.3.6 Helicity, entropy and binding affinity of constrained/truncated peptides

In this and previous studies deriving helical peptide antagonists of AP-1, helicity of antagonists and peptide–cJun complexes has been considered an important property to analyse and optimise due to the nature of CC formation (Rao *et al.*, 2013, Worrall and Mason, 2011). An early model suggested a monophasic supercoiling event between preformed fully helical structures (Crick, 1953), such that peptides with greater α-helical propensity should be more capable of supercoiling than less helical peptides (O'Neil and Degrado, 1990, Rao *et al.*, 2013).

More recent studies have pointed to an alternative biphasic model. This involves partial helix adoption by random coil monomers and concerted interaction with other monomers to form an intermediate conformation, which subsequently rearranges to form the fully folded CC (Dragan and Privalov, 2002, Mason *et al.*, 2007a). Partial helical structure is required to appropriately position side chains for interhelical enthalpic interactions to form the intermediate, and necessitates the smallest conformational change upon supercoiling (Crick, 1953, Pauling *et al.*, 1951).

However, contrary to expectations, the helicities of constrained peptides of this study correlate weakly with affinity, and high helicity alone cannot guarantee binding, as observed for JunW_{CANDI} constraint and truncation, and linear Fos and Jun derivative peptides (Rao *et al.*, 2013, Worrall and Mason, 2011, Crooks *et al.*, 2011).

Helicity affects both α-helix and CC formation processes. As discussed in 3.3.1, helix adoption reduces conformational entropy (ΔS_{conf}) which can be overcome by enthalpic backbone H-bonds, with CC formation imposing further conformational restraints and reducing ΔS_{conf} further. However, α-helix and CC formation also

increase entropically favourable desolvation (S_{desolv}) of hydrophobic groups relative to random coils or isolated α-helices respectively (Yang and Honig, 1995, Kentsis and Sosnick, 1998), such that net CC adoption entropy can be favourable if $\Delta S_{desolv} > \Delta S_{conf}$.

An increase in helical propensity of a peptide (e.g. by helix constraint) increases the helical population fraction due to effects on S_{conf}, S_{desolv} and H-bond enthalpy as discussed in 3.3.1, such that interhelical enthalpic interactions lead to CC formation. However, this is concomitant with further decreases in S_{conf} of both helix and CC formation (Zimm and Bragg, 1959, Worrall and Mason, 2011). Above a helicity threshold, ΔS_{conf} of both processes may outweigh favourable ΔS_{desolv}, and entropy opposes CC formation (Bissantz *et al.*, 2010), perhaps reflecting high peptide backbone rigidity that prefers 3.6 residues per turn (α-helices) to unwinding to 3.5 residues per turn in CCs (Crick, 1953, Pauling *et al.*, 1951). Whether helix constraint is effective in improving peptide binding relative to linear peptides thus may depend on whether the increased favourability of forming α-helices generates a lower net ΔG cost to CC formation than the further loss of S_{conf} involved in the latter process.

In summary, whilst helix constraints can improve the favourability of α-helix adoption, this may not guarantee improved favourability of CC formation. Some degree of helicity is necessary for peptides to form CCs, and so monomeric helicity is useful in estimating whether a peptide will bind, but cannot provide accurate prediction of binding or affinity. Above an as yet undetermined threshold, helicity may in fact be detrimental to CC formation. Thus, in truncating peptide antagonists, it may be more important to retain enthalpic interhelical interactions, particularly between core residues, than substantially increase helicity.

3.3.7 Conclusions

In conclusion, cJun antagonist FosW has been successfully truncated by 22 % with retention of 88 % binding free energy by the beneficial entropic pre-organisation effects of helix constraints. 29mer peptide **18**HC featuring hydrocarbon constraints demonstrated nanomolar cJun binding *in vitro*, was taken up by breast cancer cells where it localised to the cytosol and nucleus, and may potentially inhibit AP-1 once taken up. Further development of this more therapeutically attractive molecule may, in the future, generate a therapeutic agent for cancers featuring AP-1 dysregulation.

3.3.8 Future directions

In the future, further investigation of lactam and hydrocarbon stapled truncated FosW derivatives could prove valuable for the development of even more attractive AP-1 antagonists. Lactam peptide **20** required modification with cell- and nuclear-penetrating moieties to allow cellular delivery. Alternative use of hydrocarbon staples for cellular uptake of **20** unfortunately suffered from poor aqueous solubility and aggregation of FITC-**20**HC, to limit cell uptake. Further modification of **20**HC may nevertheless help to overcome these issues, supported by the fact that FITC-**18**HC was more soluble and taken up by cells. Comparing **18**HC and **20**HC, differences that could be important for cell penetration include an extra four N-terminal residues, and a higher net charge (-4 vs. -2) for **18**HC. Recent analysis of 200 hydrocarbon stapled peptides identified peptide net charge, staple position within peptides and staple type (separated vs. joined staples) to be the most influential physicochemical properties common between cell-permeable stapled peptides of the properties that were evaluated (Chu *et al.*, 2015). Cellular uptake

increased as peptide net positive charge increased, mimicking the high cell penetrance of cationic CPPs such as Tat and octoarginine. Furthermore, stapling of octoarginine improves cell penetrance, suggesting the synergy of staples and positive charge for cell uptake (Chu *et al.*, 2015). Thus, modifying **20**HC to achieve a high positive net charge could increase cellular uptake. However, peptides with net negative charge also demonstrated cellular uptake above background levels (unmodified peptides) (Chu *et al.*, 2015), corroborated by a literature survey of reported cell penetrating stapled peptides (Bird *et al.*, 2014). The increased uptake of **18**HC may also be a consequence of improved aqueous solubility relative to **20**HC, as supported by microscopy analysis reported here. Thus, increasing the magnitude of net charge of **20**HC may increase cellular uptake. To do so, ***b***, ***c*** and ***f*** positions not used for stapling, and suggested by the FosW–cJun crystal structure to participate minimally in interhelical interactions, could be mutated to charged residues. Such a derivative may then represent a more drug-like compound than longer **18**HC.

Further investigation of **18**HC and **20**HC should focus on repeating the preliminary qRT-PCR analysis of effects of these peptides on AP-1-driven *MMP9* and *Cyclin D1* expression using a larger number of biological replicates. This would enable statistically significant conclusions to be draw about AP-1 inhibition by these peptides. If AP-1 inhibition was confirmed, this could encourage testing of whether this translates to decreased breast cancer cell viability, such as through use of MTT assays, using normal human cells as controls to test for cytotoxicity other than via AP-1 activity reduction. The undesirable effects of NLS-Tat also requires further investigation; repetition of qRT-PCR analyses could include the FLU-Tat peptide to determine whether Tat, NLS or both appendages were the source of the effects

observed. Inclusion of FLU-Tat and NLS-Tat as controls for the suggested subsequent cell viability assays could then determine whether cytotoxicity of these compounds was responsible for the observed gene expression reduction.

Further modifications to **20**, **18**, **20**HC and/or **18**HC to improve the binding enthalpy of these peptides and complement the improved entropic pre-organisation achieved in this study could also be explored. Such modifications could include use of non-natural amino acids that make increased van der Waals' contacts with cJun residues at core positions, such as cyclohexylalanine substituted for Leu residues (Rao *et al.*, 2013). Determination of the structure of **20**, **18**, **20**HC and/or **18**HC with cJun by X-ray crystallography or NMR could, in addition to conclusively confirming the expected CC binding mode of these peptides, help guide which modifications to make to further improve these antagonists. Further designs that could prove fruitful if explored further include C-terminally truncated lactam **14** and **15**, as peptides not entropically pre-organised but retaining more of FosW binding enthalpy. The specificity of **20**, **18**, **20**HC and/or **18**HC for cJun over other Fos and Jun homologues is also an important parameter to ascertain, and could be achieved through use of 'dimer exchange CD' for example, where samples of peptide are mixed with samples of cJun with Fos or Jun homologues, and CD spectra monitor whether the peptide can displace Fos or Jun homologues to sequester cJun (Crooks *et al.*, 2011).

Finally, future work could focus on exploring further the relationship between helicity and affinity, enthalpy and entropy of CC formation briefly touched on in this study, using a larger dataset gathered from an extended series of peptide designs.

Chapter 4 – Novel combination of *in vitro* CIS display and *in cellulo* PCA systems to select high affinity AP-1 antagonists from diverse peptide libraries

4.1 Introduction

Peptide libraries represent a valuable source of inhibitors for disease-related PPIs such as that of oncogenic AP-1. Randomised peptides offer huge diversity in the amino acid side chain physicochemical properties, interactions, and structures that can be achieved, to afford peptides the structure and functionality to effectively modulate PPIs. Randomised library screening also has the advantage over rational antagonist design that peptide affinity for the target is not dependent on *a priori* knowledge of the most favourable peptide–target interactions (Baxter *et al.*, 2014).

A small number of direct antagonistic peptides of AP-1 have been previously described (see 1.3.2). The shortest of these, at 32 – 37 residues in length, were identified by screening of small peptide libraries (Mason *et al.*, 2006, Crooks *et al.*, 2011). However, even shorter peptides (≤≈30 amino acids) may display improved bioavailability through enhanced cellular uptake (Agrawal *et al.*, 2016) and by containing fewer protease recognition sites and immune epitopes. Thus, further downsizing of peptidic AP-1 antagonists is attractive for future therapeutic use.

To create shorter antagonists, existing peptides can be truncated; however, even with a parent molecule with high affinity for cJun such as FosW, binding affinity and structural propensity is rapidly lost with moderate truncations such that additional covalent stabilisation is essential to retain high affinity (see Chapter 3). An

alternative strategy is to optimise the primary sequence of short peptides to compensate for the loss of interaction points with the target. This goal is most easily and effectively achieved using peptide library selection systems.

To derive antagonists with favourable therapeutic-like properties as well as high affinity, *in cellulo* library selection systems have the advantage of screening for target binders within cells. This mimics the environment in which a therapeutic agent binds to its target, but suffers from the requirement for transformation/transfection of cells with library-encoding nucleic acids, limiting the library diversity that can be screened. On the other hand, *in vitro* systems select peptides under conditions which cannot necessarily replicate those within the cell. However, by avoiding cell transformation/transfection, these systems have the advantage of screening much more diverse peptide libraries such that there is much greater exploration of physicochemical, functional and structural peptide space.

Of the various *in cellulo* systems, bacterial PCA based on reactivation of mDHFR upon library-target interaction has the benefits of selecting peptides for aggregation- and protease-resistance, cytosolic stability and non-toxicity common to *in cellulo* techniques, whilst being rapid and easily performed due to the robustness of bacterial growth (Pelletier *et al.*, 1999). However, the diversities of libraries in PCA-based selection are usually restricted to $\approx\leq 10^6$ peptides due to transformation efficiencies (Mason *et al.*, 2006). Of the *in vitro* systems, CIS display (Odegrip *et al.*, 2004) combines rapid screening using highly manipulable selection pressures with coverage of similar/higher theoretical library diversities ($\leq 10^{14}$ vs. $\leq 10^{10}$ – 10^{13} peptides) (Ullman *et al.*, 2011) to mRNA, ribosome or phage display alternatives.

Previously, *in cellulo* and *in vitro* systems have only been used separately for library screening, and it was unknown whether systems could be combined synergistically. ***However, the distinct overlap between the advantages and disadvantages of these two types of system led to the hypothesis that successful combination of an in vitro and an in vivo system could exploit this overlap, to isolate high affinity peptides from hugely diverse libraries with concomitant intracellular refinement of drug-like properties.***

To this end, this Chapter describes the novel combination of CIS display and PCA ("CIS→PCA") screening systems with the aim that CIS→PCA would embody a more powerful system than either approach alone for the isolation of attractive PPI-modulatory peptides. It was also the aim of this study to attempt isolation of shorter AP-1 antagonists than previously described. Previously, PCA-derived 37mer (4.5-heptad) FosW (Mason *et al.*, 2006) was truncated to 32 residues (four heptads), and PCA was used to re-optimise six positions from a library featuring four to eight options at each position. The resultant antagonist 4hFosW displays a binding ΔG within 11 % of FosW whilst being 13 % shorter (Crooks *et al.*, 2011), such that the loss of binding affinity for the truncation achieved in 4hFosW is much less than that lost by truncation of FosW without residue re-selection (peptide "LIN20" in Chapter 3). ***This supported the hypothesis that CIS→PCA selection of more optimal residues to compensate for a smaller interaction interface could successfully identify short peptides that retain appreciable binding affinity.*** It was expected that screening a much larger library diversity than in the 4hFosW PCA ($\leq 10^{14}$ vs. 49,152 peptides) would benefit the isolation of short peptides with desirable affinities for cJun.

To complete these aims, a three-and-a-half heptad peptide library ("3.5h"), based on the FosW LZ CC domain but one heptad shorter, was designed and built using the ProxiMAX randomisation technique (Ashraf *et al.*, 2013). ProxiMAX avoids redundancy in genetic encoding of peptides and thus screening of multiple copies of identical peptides, which can bias for their selection and reduce actual diversity screened (Ashraf *et al.*, 2013). This is achieved by encoding each amino acid with a unique DNA codon, thereby removing natural amino acid incorporation bias resulting from encoding of some amino acids by multiple codons. As a result, amino acid incorporation is highly controllable, and can be used to generate semi-randomised libraries featuring particular amino acids. The library was subjected to CIS display selection against cJun and a small subset of peptides were successfully enriched, which were cloned into the PCA system with full coverage of the narrowed peptide diversity (see Figure 2.2). PCA screening rapidly isolated a single peptide named "CIS→PCAWinner" ("CPW"), which was synthesized and characterised for binding to cJun *in vitro*. CPW displayed high affinity binding to the cJun CC domain, and a novel sequence compared to FosW. These results suggest this peptide could outcompete cFos for cJun in an *in vivo* situation to antagonise cFos–cJun AP-1 formation and activity. This could make CPW attractive for further development towards a therapeutic agent, or to aid development of other peptides towards this goal.

4.1.1 Experimental Approach

The objectives of this study were as follows. Firstly, a highly diverse 3.5h library was designed. This contained amino acid options typical of CC motifs (Mason and Arndt, 2004), and some more unusual options in the form of aromatic residues

at core ***a*** and ***d*** positions and hydrophobic residues at solvent-exposed ***b***, ***c*** and ***f*** positions (see Figure 2.1 for library design). Leu residues at ***d*** positions – fixed in all other PCA screens against AP-1 to date (Mason *et al.*, 2006, Mason *et al.*, 2007b, Crooks *et al.*, 2011) – were also randomised for increased library diversity. This was hypothesized to provide increased chance for selection of more optimal residue contacts for maximal binding enthalpy to compensate for the shorter peptide length. It was anticipated that from this library a 3.5-heptad peptide of novel sequence would be selected that would demonstrate high affinity binding to cJun. The 3.5h library was constructed using the ProxiMAX system (Ashraf *et al.*, 2013) through successive library codon additions to create heptad units, which were then ligated together and PCR amplified to form the library variable region, before ligation onto CIS display DNA constructs and PCR amplification ready for CIS display (Figure 2.2 and 2.3). The constructed library was screened in CIS display against immobilised synthetic cJun CC domain peptide, at two different selection stringencies (amounts of target presented to expressed peptide library) to cover a range of possible peptide affinities for cJun. CIS display selection efficacy was monitored by deep sequencing, and peptides selected by CIS display were cloned into vectors of a modified PCA system established in *E. coli*. PCA on selected peptides consisted of Single Step and Competition Selections involving solid-phase and liquid growth respectively, using minimally nutritious M9 media and requiring library-target interaction and concomitant refolding of mDHFR for bacterial growth and peptide selection. The resultant 'winner' of CIS→PCA selection steps was analysed *in silico* for prediction of helicity and CC interaction stability, and by *in vitro* biophysical analysis following solid-phase synthesis and purification. *In vitro* biophysical characterisation of cJun

binding was performed using CD spectroscopy to assess helicity and CC binding mode, and ITC to more accurately assess interaction affinity. CC oligomeric state was analysed using size exclusion chromatography (SEC) and native polyacrylamide gel electrophoresis (native PAGE).

CD spectroscopy was performed on the winning peptide alone or a 1:1 stoichiometric mixture of peptide:cJun at 150 μM total protein concentration (P_t) in 10 mM potassium phosphate without potassium fluoride ("low salt") buffer. ITC was performed by injecting 620 μM cJun into 70 μM winning peptide, again in "low salt" buffer. SEC was used to analyse the winning peptide alone or a 1:1 stoichiometric mixture of peptide:cJun at a P_t of 150 μM in "low salt" buffer. Finally, native PAGE was used to analyse the winning peptide alone or a 1:1 stoichiometric mixture of peptide:cJun at a P_t of 600 μM and at pH 3.8 in β-alanine acetate buffer.

4.2 Results

4.2.1 CIS→PCA combination strategy

In vitro CIS display and *in cellulo* PCA library selection systems were combined for the first time to allow high throughput thorough screening of a novel library against cJun to identify AP-1 antagonists. Library construction, CIS display selection, cloning of selected peptides into PCA vectors, and subsequent PCA screening (Figure 2.2) was considered the most facile strategy for CIS→PCA combination in this proof-of-concept trial, to achieve a simple combined system using standard molecular biology and microbiology techniques, and standard laboratory equipment.

The 3.5h library variable region was constructed using the ProxiMAX technique (Ashraf *et al.*, 2013). This was attached to sequences allowing CIS display (link-RepA-CIS-Ori), which were modified from Odegrip *et al.* (2004) with an MluI restriction endonuclease site in the link region. This allowed retention of a similar, favourable C-terminal GAL helix capping motif (Doig and Baldwin, 1995) in the library peptides as the GAP motif normally used in PCA (Mason *et al.*, 2006), whilst using a standard ProxiMAX acceptor for codon addition during the library build. Ligation of complementary overhangs of AscI-digested library and MluI-digested RepA created the CIS display library construct in which AscI and MluI sites were destroyed. Following CIS display, the AscI site C-terminal of selected library peptides was reintroduced by PCR to allow NheI/AscI cloning of peptide encoding sequences into a PCA vector, which also restored the GAP C-terminal helix-capping motif. Coverage of CIS display selected peptides was guided by deep sequencing estimation of the number of particularly enriched sequences, and was easily achieved using standard PCA cloning quantities. Deep sequencing data also followed CIS display selection efficacy (peptide enrichment), and allowed evaluation of known parameters descriptive of cJun-binding capability of peptides using *in silico* prediction algorithms. Identified peptide sequences were synthesized and characterised *in vitro* to confirm cJun binding and the success of the CIS→PCA selection.

4.2.2 Truncated Library Design

4.2.2.1 Library peptide length and amino acid residues

The 3.5h library was based on re-randomisation of FosW and removal of the N-terminal heptad, such that the library C-terminus aligns with that of the 4.5-heptad

cJun CC domain. It is hypothesized that helicity correlates positively with peptide affinity for cJun (O'Neil and Degrado, 1990, Rao *et al.*, 2013), suggesting that the least helical heptad of FosW could be removed to design a truncated peptide library without substantial loss of affinity. The Agadir algorithm predicts that the N-terminal heptad of FosW is poorly helical (Figure 4.1) (Lacroix *et al.*, 1998), making this heptad a good candidate for truncation (Crooks *et al.*, 2011).

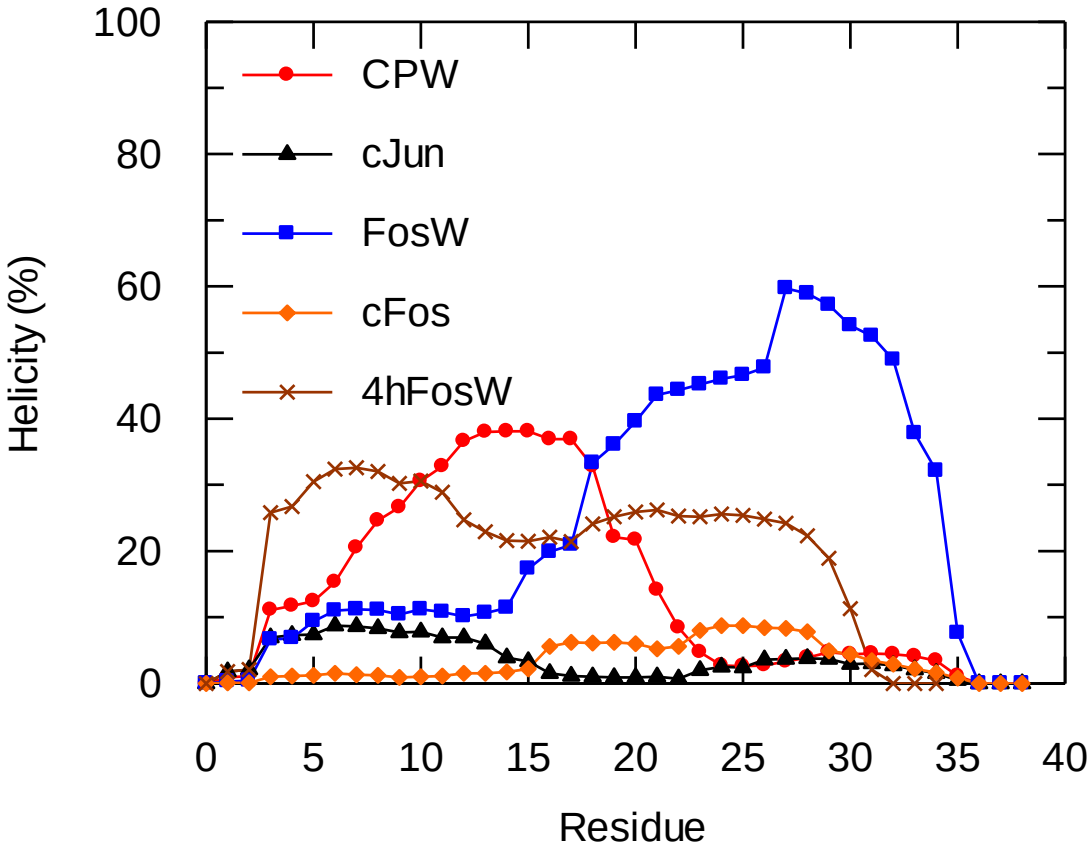


Figure 4.1: Per residue helicity of CPW compared to that of cJun, FosW, cFos and 4hFosW. Helicity on a per residue basis, as calculated by Agadir (Lacroix *et al.*, 1998).

Library amino acid options were rationally chosen based on residues known to stabilise parallel dimeric CC interactions (Mason and Arndt, 2004) opposite residues of cJun, and based on observed frequencies in 197 natural LZ CCs (Crooks, 2013). This should supply options for high interaction affinity and a strong preference

for peptides to adopt CC structures with cJun over other binding modes. Finally, more unusual options from the perspective of residue frequencies and known interactions were included to allow the chance for beneficial serendipitous selection.

The theoretical diversity of the 3.5h library was 7.3 x 10^{17} different peptides. This did exceed the theoretical screening capacity of CIS display (≈10^{14} peptides) (Ullman *et al.*, 2011). However, in this proof-of-principle of CIS→PCA, the primary focus was simply to screen a library diverse enough to contain cJun binding peptides, such that partial coverage of the library was satisfactory.

4.2.2.2 Core *a* and *d* residue options

At core ***a*** positions, cJun consists of near-ideal residues for a hydrophobic parallel dimeric CC core (see Figure 4.7B), the only exceptions being ***a'***3 and ***a'***4 residues occupied by a polar, parallel- and dimer-specifying Asn residue and a helicity-inducing Ala residue respectively (Oakley and Kim, 1998, Lumb and Kim, 1995, O'Neil and Degrado, 1990). At core ***d*** positions, cJun exclusively consists of dimer-specifying hydrophobic Leu residues (Harbury *et al.*, 1993).

Opposite these cJun residues, ***a*** and ***d*** position library options included Ile, Leu, and Val (all included at equal frequency, except for aromatic residues [see below]). The most favourable residue for maximal dimerisation ΔG from ***a–a'*** and ***d–d'*** interactions with cJun core residues would be Ile (Acharya *et al.*, 2006), and so Ile would be expected to be a favourable selection. The parallel packing preference of side chains at ***a*** positions should also favour Ile over hydrophobic alternatives Leu and Val, though Val with its smaller side chain may be selected if space is limited for Ile packing (Harbury *et al.*, 1993). Conversely, the perpendicular packing preference

at ***d*** positions favours Leu selection in LZs (Harbury *et al.*, 1993, Crooks, 2013), despite Ile–Leu providing an extra -0.5 kcal/mol/dimer to ΔG (Acharya *et al.*, 2006). Thus, whether Ile or Leu is selected at ***d*** positions may be determined by necessity for higher enthalpy (Ile) or preference for perpendicular packing (Leu). Val, as for Ile, does not favour perpendicular packing, but may be selected for its small size.

Asn was also introduced at the Heptad 2 ***a*** position to provide opportunity for N–N homotypic interaction with the ***a'***3 N of cJun. This interaction is more thermodynamically favourable than hydrophobe–Asn interactions and so ***a2*** Asn selection would be expected (Acharya *et al.*, 2006). Asn was also included at ***a3*** to simplify ProxiMAX library construction through use of identical Heptad 2 and Heptad 3 units at the heptad ligation stage (see Figure 2.3). However, selection of Ile at this position should provide the greatest gain to dimerisation ΔG (Acharya *et al.*, 2006).

Aromatic Phe, Trp and Tyr residues were also included at all core positions at a total frequency of 5 % (equal representation of each residue). Though observation frequencies of aromatics are low at core positions, they do display unexpected over-representation at ***a*** positions (Crooks, 2013). Because shorter antagonists may require interactions of increased enthalpy to compensate for a loss of interaction points, inclusion of bulky aromatic side chains provided opportunities for contacts of higher enthalpy than Leu/Ile/Val. Aromatics may also confer heterodimer preference through disfavoured self-interaction (Smeal *et al.*, 1989). Inclusion of these residues additionally provided an opportunity to further probe their function at these positions. However, aromatic selection is only likely where enough room is available in the core for their large side chains. Analysis of the FosW–cJun crystal structure

(Chapter 3), assumed to represent a well packed dimeric core, indicates that a few of the commonly observed aromatic side chain rotamers (Schrodinger, 2013) could fit in the 'holes' generated by surrounding cJun side chains with small rearrangements. The most likely positions for aromatic selection are the library ***a***4 and ***a***4.5 positions, aided by the small A and V side chains at ***a'***4 and ***a'***4.5 respectively, and at ***d***4.5, aided by the absence of ***e'***4.5 and ***a'***5 side chains due to proximity to the C-terminus. Aromatics were nonetheless included at all core positions to simplify library construction, and provide flexibility and chance for serendipity in their selection.

4.2.2.3 Core-flanking *e* and *g* residue options

At ***e*** and ***g*** positions, cJun has E/K/A/R and K/Q/T/Q residues (heptads 1/2/3/4) respectively. To provide opportunity for attractive, stabilising interhelical electrostatic interactions, one negatively charged (Glu) and two positively charged residues (Lys and Arg) were included as library options. cJun Glu residues would be best paired with Arg, whilst Lys would be best paired with Glu. In the case of cJun Ala and Gln, the most preferable amino acid is another Gln (Krylov *et al.*, 1998), and so this was included as the fourth option. These four amino acids represent the most frequent and most over-represented residues at ***e*** and ***g*** positions (Crooks, 2013). It was expected that charged residues would be selected where favourable $\boldsymbol{e}_{i}$–$\boldsymbol{g'}_{i-7}$ and $\boldsymbol{g}_{i}$–$\boldsymbol{e'}_{i+7}$ interactions were possible, and Gln selected otherwise. Asp was not included as an option because Glu has both higher helical propensity (Padmanabhan *et al.*, 1996), and observation frequency in CC ***e*** and ***g*** positions. The breadth of residues included should also provide opportunity for selection of favourable interhelical ***a***–***g'*** and ***e***–***d'*** interactions, which are known to be involved in CC stabilisation (O'Shea *et al.*, 1991, Glover and Harrison, 1995, Havranek and Harbury, 2003).

4.2.2.4 Solvent-exposed *b*, *c* and *f* outerface residue options

Library options at outerface positions were not expected to be able to directly interact with cJun residues, and so were chosen for possible involvement in intrahelical interactions, or solubility-/helicity- promotion. Residues A, E, I, K, L, Q and R thus span potential for intrahelical interactions, solubility (Glu≈Lys≈Arg>Gln) or helicity induction (Ala>Arg≈Lys≈Leu>Gln≈Glu>Ile) (O'Neil and Degrado, 1990). Additionally, hydrophobic Ile and Leu were included to probe the ill-understood preferences of ***b***, ***c*** and ***f*** residues further. It was expected that outerface residues would be chosen in a position-specific manner to maximise peptide competency for dimerisation, dependent on the physicochemical properties and spatial orientations of surrounding amino acids (Grigoryan *et al.*, 2009, Crooks, 2013). At outerface positions, A, E, K, Q and R are in the top ten most frequently observed and most enriched residues (Crooks, 2013), making good library options. Though unusual choices, Ile and Leu are also found at ***b***, ***c*** and ***f*** positions (Crooks, 2013), where they may form *i→i+4* van der Waals' interactions with polar/charged residue side chains, and/or contribute to helicity (Andrew *et al.*, 2001, O'Neil and Degrado, 1990).

4.2.2.5 Helical residue options

Finally, library options all score well for helical propensity, and were chosen with the expectation that peptides need to be of appreciable helicity for supercoiling (O'Neil and Degrado, 1990, Rao *et al.*, 2013). Infrequent cases where other residues were more enriched than those chosen as library options, such as Ser vs. Arg/Lys at ***b*** and ***c*** positions, were also generally of lower helicity than options chosen and so were excluded. Cys is seemingly more enriched than Ile, Leu or Tyr at ***a*** positions, and has a higher helical propensity than I, V, F, W and Y (O'Neil and Degrado, 1990). This

is, however, likely to be an anomaly caused by a very low overall frequency in the LZ dataset (Crooks, 2013). Methionine is also more enriched than V, F, W or Y at ***d*** positions, though again this argued to be an anomaly caused by the huge over-representation of Leu at ***d*** positions. Additionally, Cys was not included due to potential to form undesirable disulfide bonded peptide homodimers which could result in removal of otherwise valuable cJun binders from the selection. Finally, peptide helicity was further encouraged through use of AS and GAP helix-capping motifs to library peptides (Doig and Baldwin, 1995).

4.2.3 Library construction, QC and diversity

Cycles of successful library codon ligation to Heptad 1, 2 and 4.5 acceptors, and MlyI digestion ready for addition of the next library codon are shown for construction of Heptad 1, 2 and 4.5 cassettes in Appendix Figure 4.1, 4.2 and 4.3. Generation of the desired products was evidenced by observation of bands of the expected molecular weights on agarose gels. Completed Heptad cassettes were then ligated together and PCR amplified, and the completed 3.5h variable region (Heptad 1-2-2-4.5) was deep sequenced to assess library quality. From a random and thus representative sample of 5,038,153 sequence reads, observed codon incorporation frequencies for peptides of the expected length (Figure 4.2) were very close to those expected from the library design (Figure 2.1), and the orientation of Heptad cassettes was also as expected, suggesting that peptides of the correct randomisations had been successfully constructed. Following this, appendage of the RepA CIS display construct to the 3.5h library variable region was performed, and Sanger sequencing confirmed the expected RepA sequence (proprietary, data not shown).

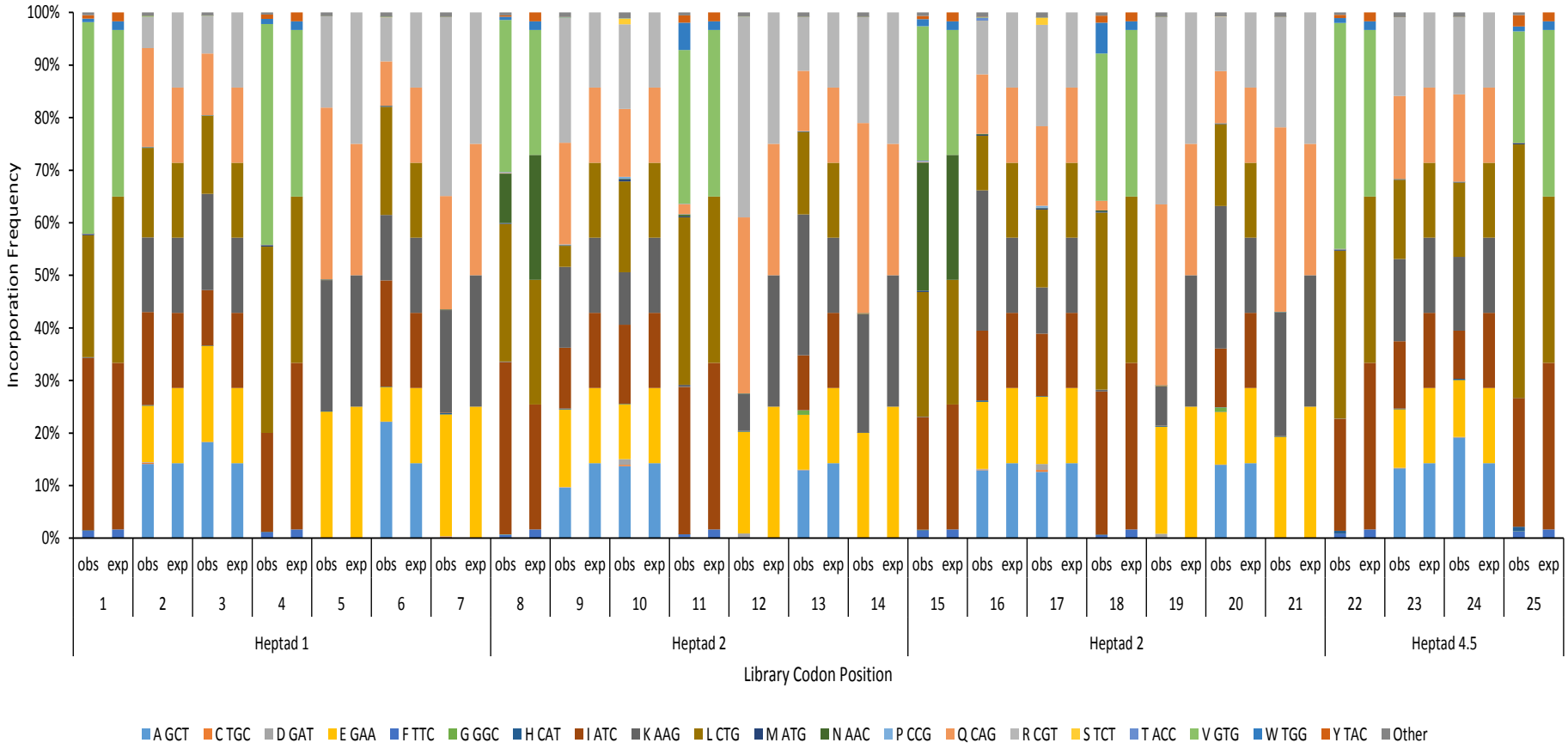


Figure 4.2: Deep sequencing QC of codon incorporation frequencies observed vs. expected for the completed 3.5h library variable region. Observed (obs) and expected (exp) incorporation frequencies are shown for library amino acids/codons (coloured) at positions 1 – 25 of correct length peptides of the completed 3.5h variable region (Heptad 1-2-2-4.5).

However, deep sequencing also identified peptides having a length other than the designed 25 amino acids. Sequences of the correct length made up 39 % of the library, with longer constructs making up 17 % of the library, and shorter constructs 44 %. Shorter constructs can be attributed in part to single nucleotide and dinucleotide deletions (5 % each), a missing codon (4 %) or a missing heptad (5 %), whilst the remaining ≈24 % are a variety of other shorter constructs.

The sources of these incorrect length peptides are likely due to errors in endonuclease digestion, PCR fidelity and primer annealing, and in oligonucleotide reagents during codon addition and heptad ligation stages, whilst an expected MiSeq error rate of <1 % (Laehnemann *et al.*, 2016) suggests that few shorter/longer peptides are the result of sequencing errors. PCR amplification is probably the largest source of error: despite use of Phusion high fidelity polymerase with a reported error rate of 4.4×10^{-7} bp^{-1} (ThermoScientific, 2016b), multiple rounds of PCR for codon addition and then PCR of ligated heptads, using pmol quantities of DNA as PCR template ($>6 \times 10^{11}$ molecules), undoubtedly results in significant error, even before compounding with other errors. Previously, it has been reported that ≈25 % of a 21mer peptide library built by ProxiMAX was shorter or longer than the library design (Frigotto *et al.*, 2015). The increased fraction of the 3.5h library displaying length errors could be due to construction of a longer peptide library via a greater number of codon additions and total PCR cycles than the study by Frigotto *et al.* (2015).

Peptides with a missing residue (single codon deletion) represent the least problematic of the shorter peptides observed, as these will still be in-frame and so likely functional. Peptides with single/dinucleotide deletions conversely will result in

frameshifts, such that some peptides may retain function whilst others do not, depending on exactly which amino acids are encoded. Substantially shorter peptides are likely to have lower affinities than correct length peptides, such that correct (or near-correct) length peptide selection is favoured. Peptides longer than expected, on the other hand, could be preferentially selected over desired length peptides through formation of more points of interaction with the target (and thus an increased interaction affinity). However, peptides of the correct length or that are shorter than desired outnumber longer peptides (82.5 % of the library vs. 17.5 %) such that the likelihood of selection of the former two groups of peptides was favourable. Furthermore, it is possible that more optimal residue choices in short peptides may provide greater benefit to peptide affinity than an increased number of interaction points with the target, as was the hypothesis of this study, though it is noted that more optimal residue selections are equally likely in peptides of any length.

The actual library diversity presented to cJun in CIS display selection was calculated to be ≈6.1 x 10^{10} peptides (see section 2.5 for calculation). Of this diversity, the issues with peptide length meant that 39 % of peptides screened (i.e. 2.4 x 10^{10}) were of the correct length. It is worth noting that whilst the diversity screened is substantially less than the theoretical diversity (7.3 x 10^{17}), screening theoretical diversities is an issue for any library selection system due to the prohibitive quantities of DNA required. Furthermore, deep sequencing of the 3.5h library observed each peptide (of any length) only once, indicating that library diversity screened by CIS→PCA was maximised through avoiding over-representation of particular peptides, a problem that other library construction techniques suffer from and a major benefit of ProxiMAX (Ashraf *et al.*, 2013). Finally, the 3.5h library still

represented a greater searching of sequence space than previous screening against cJun (Mason *et al.*, 2006, Crooks *et al.*, 2011), and so was expected to yield novel peptides which may prove valuable for AP-1 antagonism.

4.2.4 CIS→PCA selection of "CIS→PCAWinner" peptide from the 3.5h library

4.2.4.1 Efficiency of CIS display selection and peptide enrichment

The success of CIS→PCA was dependent on the selection pressures in CIS display and PCA being compatible with each other, such that peptides selected *in vitro* could subsequently bind cJun under different conditions *in cellulo*. Previously, CIS display of a library doped with FosW was able to preferentially select this sequence (Rao, T., 2010, unpublished results), demonstrating that CIS display could select high affinity cJun binders. By applying the same selection conditions in the selections performed here on the 3.5h library, it was hypothesized that CIS→PCA could be successful in selecting cJun binders. These conditions related to the strength ("stringency") of the selection pressure applied for specific binding to the target, the affinity of which can be increased by decreasing the quantity of cJun coated on magnetic beads during CIS display selection. Two selection stringencies were employed: a "low stringency" (higher cJun quantity), and a "high stringency" (lower cJun quantity) (for further discussion, see 4.3.1).

The progress of the selection was monitored by deep sequencing of selected peptide DNA after selection rounds R3 and R4, which identified 25,000 – 72,000 unique sequence reads (depending on sample). From these, 85 – 131 peptides were observed ≥50 times and considered to have been 'enriched', warranting further analysis. In the low stringency selection, at selection rounds R3 and R4 none of the

enriched peptides were the correct length and 65 % were 32mers (4.5-heptads long), whilst in the high stringency selection 27 % of peptides were the correct length and 11 % were 32mers, such that in both conditions the majority of selected peptides were not of the expected length. The top 20 peptides are ranked by observation frequency in Appendix Table 4.1, and the overall composition of the remaining lower ranked peptides sequenced are given in Appendix Table 4.2.

Observation frequencies for peptides observed ≥50 times are plotted in Figure 4.3, and displayed with the highest ranked peptide sequences in Table 4.1. In R3 of the low stringency selection (Figure 4.3A), three sequences were successfully enriched with relative observation frequencies of 19 %, 8.2 % and 7.9 %, whilst the remaining 97 % of sequences each had relative frequencies ≤5 %. The enrichment of these three sequences continued in R4 as would be expected for an efficient selection, where relative frequencies increased by 1.2-fold – 1.6-fold vs. R3. Thus, the low stringency CIS selection appears to have positively enriched peptides.

In R3 of the higher stringency selection, Figure 4.3C indicates lesser enrichment than the low stringency condition. Again, three peptide sequences were particularly enriched, but their relative frequencies were much lower than the equivalent selection round of the low stringency selection. However, the most enriched peptides in R4 (Figure 4.3D) do reach similar frequencies to the peptides enriched in R4 of the low stringency selection, suggesting that the high stringency selection may be enriching peptides.

The number of unique peptides enriched after R3 of the high stringency selection increases slightly by R4, contrary to expectations (though it is noted that

this increase is slight and is based on a small sample). If this is an accurate representation of the selection, this suggests that non-specific binders are not outcompeted for the target by a smaller number of specific binders, such that further 'unique' peptides are created from these peptides by errors in PCR amplification. Conversely, there is a slight decrease in the number of unique sequences from R3 to R4 for the low stringency selection (again with the same caveats) which may indicate that the low stringency selection was enriching peptides more efficiently. Collectively, the above analyses suggest that the low stringency selection was the most likely to deliver valuable target binders in subsequent PCA.

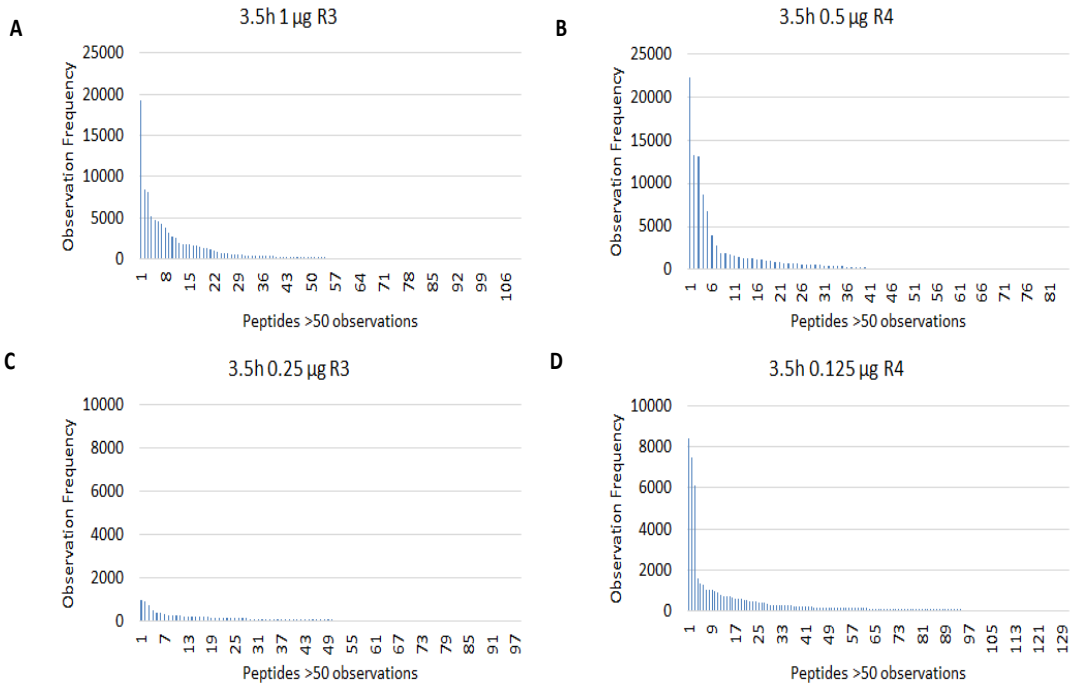


Figure 4.3: CIS display low stringency selection has more successfully enriched distinct peptide sequences from the 3.5h library. Observation frequencies of peptides observed ≥50 times in the peptide subset selected by CIS display, as identified by deep sequencing. Peptides ranked in order of observation frequency. **A** and **B**; peptides from the 3.5h low stringency selection round R3 and R4. **C** and **D**; peptides from the 3.5h high stringency R3 and R4. cJun quantities determining selection stringency are given in µg (see Table 2.11 for further explanation).

Table 4.1: Success of the low stringency CIS display selection in enriching particular sequences predicted to be cJun binders, and favourable comparison of subsequent CIS→PCA low stringency selection winner CPW to these enriched peptides. Displayed are peptide observation frequencies, and relative observation frequencies as a percentage of all observed sequences, from deep sequencing of CIS display low and high stringency selection R3 or R4 outputs, and for subsequent winner of the CIS→PCA low stringency selection CPW (bold text) and comparison peptide CIS21. Peptide helicities were calculated by the Agadir algorithm (Lacroix *et al.*, 1998), predicted interaction T_m values were calculated by the bCIPA algorithm (Mason *et al.*, 2006) and predicted interaction stability by the BOW algorithm (Fong *et al.*, 2004), and predicted aggregation propensities were calculated using PASTA2.0 (Trovato *et al.*, 2006) and Zyggregator (Pawar *et al.*, 2005) algorithms. *AS and GAP helix capping motifs added for Agadir, PASTA 2.0 and Zyggregator analyses. †Where two values are given for bCIPA and BOW scores, this relates respectively to N-terminal and C-terminal alignment of the peptide to its longer cJun partner; the more positive bCIPA and BOW score indicates the preferred binding site. ‡Too short for BOW analysis (28 – 60 amino acids required). Values are given to 2 s.f. except observation frequency (raw number). Low predicted aggregation propensity is indicated by a more positive PASTA 2.0 score, or a more negative Zyggregator score, and via comparison with known soluble peptides cFos (-3.0/-6.8), FosW (-1.5/-7.3) and cJun (-1.7/-5.5) (PASTA 2.0/Zyggregator scores respectively).

Selection stringency	Position in CIS selection	Peptide name	Peptide sequence	Observation frequency	Relative observation frequency (%)	Helicity (%) (Agadir)*	Predicted interaction T_m (°C, bCIPA)	Predicted interaction stability score (BOW)	Predicted aggregation propensity score (PASTA 2.0/Zyggregator algorithms)*
3.5h CIS low stringency	1st in 1 µg R3		IRKLELEIEAIEAELLEIRAQLEEIQLQVIQL	19200	19	22	44	28	-3.8/-4.2
	2nd	CIS1	VQEIELQLEELEKQNKLIQEQLIAVQKEIEQV	8440	8.2	5.4	44	16	-4.2/-4.2
	3rd		IAKLEKRIIELQIKLLQAQVEKVQLEIKAVEEQVRLV	8130	7.9	27	17/33†	8.8/22†	-4.2/-2.8
	1st in 0.5 µg R4	CIS1	VQEIELQLEELEKQNKLIQEQLIAVQKEIEQV	22200	22	5.4	44	16	-4.2/-4.2
	2nd		IKLVQREIAAWRIKNQQLRIEIRQLEAI	13200	13	4.1	7/-52†	27/0.80†	-3.4/-3.4
	3rd		IRKLELEIEAIEAELLEIRAQLEEIQLQVIQL	13100	13	22	44	28	-3.8/-4.2
3.5h CIS high stringency	1st in 0.25 µg R3		VRIVELQIVRIQNARIQLELKRVRQRILAL	1010	7.0	9.7	-15/3†	-3.2/15†	-6.1/-4.0
	2nd		IEIIKQQIIIVQARIRAIQLQVKLVRKQVLQV	913	6.3	6.3	25	7.7	-8.4/-3.7
	3rd		IKKIELQLLEKRIQIVRLKIQI	781	5.4	10	-17/-55†	‡	-5.2/-3.7
	1st in 0.125 µg R4		VRIVELQIVRIQNARIQLELKRVRQRILAL	8390	16	9.7	-15/-74†	-3.2/15†	-6.1/-4.0
	2nd		IEIIKQQIIIVQARIRAIQLQVKLVRKQVLQV	7460	14	6.3	25	7.7	-8.4/-3.6
	3rd		IKKIELQLLEKRIQIVRLKIQI	6110	11	10	-17/-55†	‡	-5.2/-3.7
3.5h CIS→PCA low stringency	**71st**	**CPW**	**VQEIEQEIQELEKRIKQIQQEFQEIEQQIALL**	**74**	**0.1**	**15**	**53**	**24**	**-3.6/-7.0**
3.5h CIS low stringency	21st	CIS21	VQKLELEIEQLEQELLLIQAQLKKVELELKRL	872	0.9	46	49	26	-4.2/-3.3

4.2.4.2 CIS display selected peptides

The top three peptides in each stringency selection (Table 4.1) represent the most enriched peptides after CIS display, and so could represent the highest affinity binders. Indeed, some of these peptides have favourable predicted interaction capabilities with cJun (see Table 4.1 and 4.2.5.2), indicating the low stringency selection in particular was enriching promising cJun binders. However, the top ranked peptides are not necessarily those of highest affinity; peptides with affinity close to/higher than the threshold determined by the selection stringency (target quantity) can be selected in a near quantitative fashion, leading to a more arbitrary ranking (Jalali-Yazdi *et al.*, 2016). Furthermore, *in vitro* selection can feature secondary selection pressures that can enrich peptides binding non-specifically to other components of the selection such as magnetic beads, streptavidin, biotin or polypropylene plates (Vodnik *et al.*, 2011). As a result, high affinity, high specificity cJun binders that could win the subsequent PCA may be present in the CIS display enriched subset but ranked lower than other peptides, which may or may not represent less specific target binders. Thus, unique peptides selected by CIS display and identified by sequencing were cloned into PCA vectors to allow any peptide to be selected in PCA regardless of CIS display ranking. As it transpired, the most enriched peptides from the low and high stringency CIS display selections were not the winners of the corresponding PCAs, indicating the importance of transferring all CIS display selected peptides to PCA.

Of those peptides observed ≥50 times, in the positively enriching low stringency selection none were the correct length, and instead 32mers (4.5-heptads in length, the same length as cJun), constituted 65 % of both the R3 and R4 pools.

Conversely, in the less efficiently enriching high stringency selection, 27 % of peptides at R3 and R4 were the correct length. This could support the hypothesis that longer peptides were generally capable of outcompeting correct length peptides through a greater number of interactions with cJun.

4.2.4.3 PCA Selection

Deep sequencing after CIS display selection round R4 identified 25,000 – 72,000 unique sequence reads of selected peptides for transfer to PCA. Enough DNA to ensure near-complete coverage of this peptide diversity was used for PCR to reintroduce an AscI restriction site 3' of the library peptides, ligation of NheI/AscI-digested amplicons into pET28a+, XL-1 Blue amplification of ligated plasmids and PCA initiation by electroporation of library plasmids into PCA cells. The diversity of these peptides was covered to between 95 (high stringency) and >99 % (low stringency) in the corresponding PCAs according to Equation 1 (see 2.4.3), allowing adequate chance for *in cellulo* selection of cJun binders from the CIS enriched peptide subset.

The PCA performed here (see 2.7) was a slightly modified PCA selection to that described previously (Pelletier *et al.*, 1999), which used fewer plasmids to host library, cJun and *lacI*^{q} expression control allele sequences. This enabled use of fewer antibiotics for selective culturing, which should lessen unnecessary stress on PCA cells. PCA Single Step selection resulted in successful growth of colonies on M9 minimally nutritious agar, and selected peptides were identified by selective Sanger sequencing of extracted pET28a+ plasmids. Colonies were then scraped into M9 liquid media for PCA Competition Selection, the progress of which was monitored at each passage by sequencing three colonies and the passage pool.

For the low stringency 3.5h CIS output, the first four colonies to appear on the corresponding PCA plate and be sequenced all contained the same sequence. Because the fastest growing colonies host peptides with the highest affinity for cJun (Pelletier *et al.*, 1999), this peptide (named "CIS→PCAWinner" or "CPW" for short) should represent the overall winner of CIS→PCA for the low stringency selection.

Usually Competition Selection is used to amplify subtle differences between peptides with similar affinities to identify the highest affinity binder (Pelletier *et al.*, 1999). In the case of CPW selection by Single Step selection, this peptide had already been identified as the highest affinity binder. However, multiple nucleotides at some base positions were identified at low levels in sequencing chromatograms of the Single Step selection colonies. Thus, to rule out the possibility of other binders, Competition Selection was performed. This further provided opportunity for infrequent occurrence of mutations in peptide encoding DNA as a result of bacterial host replication errors to identify a mutant of CPW with even higher affinity. Sequence chromatograms of P2 – P5 Competition selection pools are displayed in Figure 4.4. In the P2 pool chromatogram, there are multiple positions (particularly towards the C-terminus) where two nucleotides are competing for the base call, in many cases separated by a very marginal difference in fluorescence intensity. By passage P3, the pool showed insignificant levels of secondary nucleotide signals, leaving the single clean sequence of CPW to be confirmed in P4 and P5 pools.

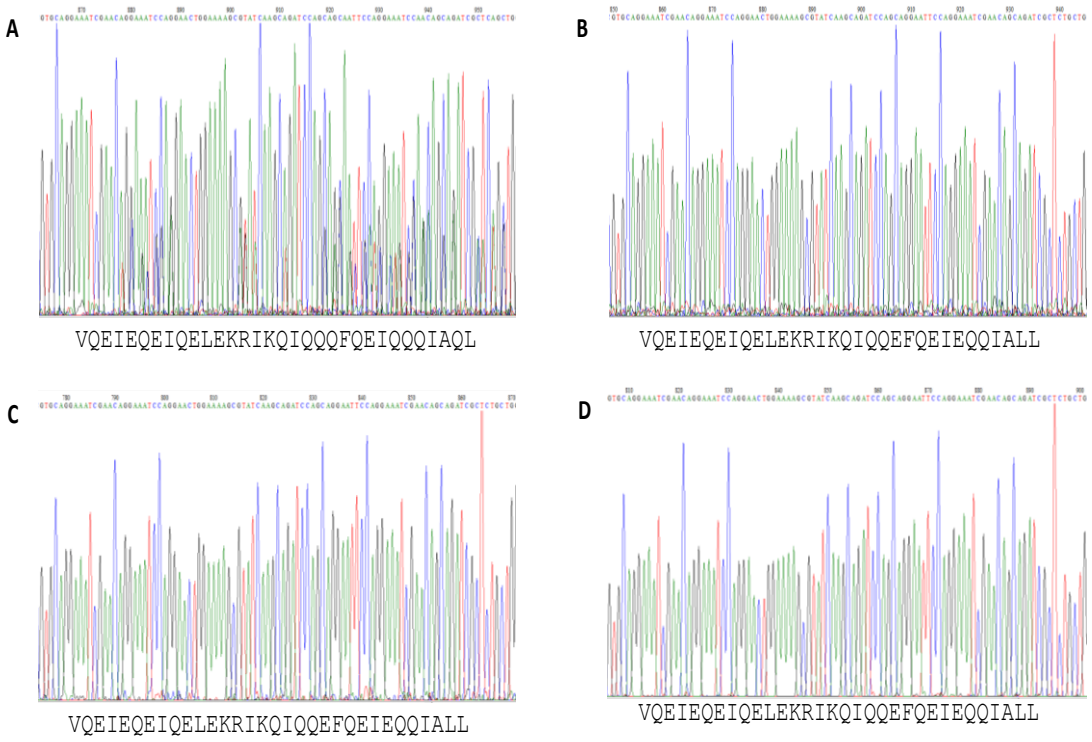


Figure 4.4: Peptide CPW is the clear winner of PCA following CIS display low stringency selection. PCA passage sequencing chromatograms. **A** 3.5h 0.5 µg P2 pool **B** 3.5h 0.5 µg P3 pool **C** 3.5h 0.5 µg P4 pool **D** 3.5h 0.5 µg P5 pool. Relative fluorescence intensities are shown per base sequenced.

CPW was thus confirmed by Single Step and Competition Selection to be the winner of CIS→PCA low stringency selection. This peptide outcompeted ≈25,000 other peptides transferred from CIS display to PCA, rising from position 71 in the low stringency CIS display output based on observation frequency.

Interestingly, for the high stringency CIS→PCA selection, the peptide ranked 1st by CIS selection afforded fastest colony growth to win the corresponding Competition Selection (named "CIS1"). This peptide was also selected and ranked 1st by the low stringency CIS selection but was ultimately outcompeted by CPW in the following PCA. These peptides are compared in the following analyses.

4.2.5 CPW, a novel, high affinity cJun antagonist

4.2.5.1 *In silico* prediction of cJun binding

In silico algorithms were used to predict the likelihood of CPW–cJun interaction, and whether this interaction is a CC interaction (as expected from the library design) rather than an alternative binding mode. These analyses (Table 4.2) corroborate biophysical *in vitro* confirmation of CPW–cJun interaction, whilst providing information not easily gained *in vitro* (e.g. CPW monomeric helicity). Furthermore, predictions by these algorithms are scaled to measured interactions, and thus provide a more representative estimate of interaction capability than manual comparison of pairwise interaction preferences. *In silico* algorithms used were Agadir (Lacroix *et al.*, 1998), bCIPA (Mason *et al.*, 2006), and BOW (Fong *et al.*, 2004). CPW's propensity for cJun binding according to these algorithms is compared with other known cJun binders (see Table 4.2): native binders cFos, Fos homologues FosB, Fra1 and Fra2, Jun homologues JunB and JunD, and cJun itself (Angel and Karin, 1991); and PCA-derived cJun binders FosW, 4hFosW (Crooks *et al.*, 2011) and $FosW_{CANDI}$ (Mason *et al.*, 2007b). CPW's propensity for binding cJun is also compared against other CIS display enriched sequences (Table 4.1).

Table 4.2: Predicted helicity of CPW, and stability/probability of interaction with cJun is comparable to previously derived cJun antagonists, and higher than native Fos and Jun proteins. Helicity calculated by Agadir (Lacroix *et al.*, 1998), predicted T_m of CPW–cJun interaction calculated by bCIPA (Mason *et al.*, 2006), and predicted stability of CPW–cJun interaction calculated by BOW (Fong *et al.*, 2004). *AS and GAP helix capping motifs added for Agadir analysis. †Peptide sequence selected by PCA (Mason *et al.*, 2006). §Peptide sequence selected by PCA (Crooks *et al.*, 2011). ¶Peptide sequence selected by PCA (Mason *et al.*, 2007b). ‡Sequences of native cJun binders from Angel and Karin (1991). 4hFosW has two values for bCIPA and BOW scores, which relate respectively to N-terminal and C-terminal alignment of 4hFosW to its longer cJun partner; the higher bCIPA and BOW scores for N-terminal alignment indicate this as the preferred binding site. Values are given to 2 s.f..

Selection	Peptide Name	Peptide Sequence	Helicity (%) (Agadir)*	Predicted Interaction T_m (°C, bCIPA)	Predicted Interaction Stability Score (BOW)
3.5h CIS→PCA low stringency selection	CPW	VQEIEQEIQELEKRIKQIQQEFQEIEQQIALL	15	53	24
PCA†	FosW	LDELQAEIEQLEERNYALRKEIEDLQKQLEKL	26	63	41
PCA§	4hFosW	LDELQREIEQLEELNYALQKEIEDLQKQ	22	32/-45	32/15
PCA¶	$FosW_{CANDI}$	LDELQAEIEQLEDQNYALQKEVEDLRKELEKL	22	57	39
Native	cFos‡	TDTLQAETDQLEDEKYALQTEIANLLKEKEKL	3.5	13	27
Native	FosB‡	TDRLQAETDQLEEEKYELESEIAELQKEKERL	5.0	14	27
Native	Fra1‡	TDFLQAETDKLEDEKYGLQREIEELQKQKERL	14	12	26
Native	Fra2‡	TEKLQAETEELEEEKYGLQKEIAELQKEKEKL	4.1	18	25
Native	cJun‡	IARLEEKVKTLKAQNYELASTANMLREQVAQL	3.7	7.0	18
Native	JunB‡	IARLEDKVKTLKAENYGLSSTAGLLREQVAQL	2.4	-4.0	19
Native	JunD‡	ISRLEEKVKTLKSQNYELASTASLLREQVAQL	2.6	2.0	18

4.2.5.2 CPW helicity and interaction stability vs. previous cJun binders

Agadir reports the helical propensity of monomeric peptides averaged over all residues (Table 4.2) and per residue (Figure 4.1) (Lacroix *et al.*, 1998). Supercoiling capability has a complex dependence on peptide helicity (see 3.3.6), but helicity is important to analyse and optimise as peptides must be at least partially helical to

bind. The average helicity of CPW (15 %) is less than the helicities of high affinity cJun binders FosW, 4hFosW and $FosW_{CANDI}$, but higher than all native cJun binders, suggesting CPW could bind cJun with increased affinity vs. native binders.

Analysis of per residue helicity (Figure 4.1) indicates that CPW's helicity is concentrated towards its N-terminus, with C-terminal helicity similar to that of cJun. This is unlike the helicities of cFos and FosW which are concentrated towards their C-termini, and thus have been argued to complement the higher N-terminal helicity of cJun for improved helicity along the length of the CC (Mason *et al.*, 2007a). Instead, the high N-terminal helicity in CPW is akin to the high N-terminal helicity of 4hFosW (Crooks, 2013) and successful lactam helix-constrained truncated FosW derivatives **14** and **15** from Chapter 3. Assuming CPW binds to cJun, N-terminal helicity in CPW may represent selection of helicity-promoting residues that, in conjunction with cJun helicity-promoting residues, create a helical site for CC nucleation (Frank *et al.*, 2000).

bCIPA and BOW algorithms were used to predict interaction stability of CPW with cJun, outputting a T_m or interaction stability score (more positive scores indicating more favourable interaction) respectively. Compared to other PCA-derived sequences (Table 4.2) – whose *in vitro* measured T_m values are in close agreement to those predicted by bCIPA – CPW's interaction T_m with cJun (53 °C) indicates promising potential for high affinity binding. Further, this T_m is far higher than all native cJun binders, the highest of which is Fra2 at 18 °C. BOW likewise predicts CPW to interact favourably with cJun, though to a lesser degree than bCIPA. The BOW score for CPW–cJun (24) is somewhat lower than PCA-derived cJun antagonists (32 – 41), but is

comparable to native Fos homologues (25 – 27), and is higher than native Jun homologues (18 – 19), to suggest CPW–cJun complexes could form favourably.

4.2.5.3 CPW helicity and interaction stability vs. other enriched peptides

CPW was selected by the CIS→PCA low stringency selection over other peptides more enriched by CIS display. Analysis of the helicity of all enriched peptides from this selection indicates that CPW's helicity is half that of the average helicity (30 ± 19 % (mean ± SD)) of peptides at R3 or R4 (Appendix Figure 4.4). Even in the top 25 enriched sequences, there are numerous peptides with higher helicity than CPW; for example, the peptide observed at position 21, hereafter referred to as "CIS21" (chosen for other interesting properties, see below), has a helicity ≈3-fold higher than CPW. Comparison against the top three sequences including the most enriched peptide CIS1, which were all outcompeted by CPW in the subsequent PCA, indicates that CPW is the third most helical sequence (Table 4.1). Thus, CPW appears not to have been selected over other peptides for increased helicity. This does not necessarily translate to reduced CPW affinity relative to other peptides: binders such as 4hFosW with low predicted monomeric helicities similar to that of CPW (Table 4.2) nevertheless form CCs with cJun of high helicity and stability (Table 4.3).

bCIPA and BOW analysis of all CIS enriched peptides was not performed due to inability to simultaneously input multiple sequences into these algorithms. Instead, the top ranked peptides from each selection stringency, and CIS21 which displayed an attractive Agadir helicity (Table 4.1), were analysed. Relative to these peptides, CPW has the highest bCIPA T_m and third highest BOW score, with a bCIPA T_m just slightly higher than CIS21, though the BOW score is slightly lower than CIS21.

This suggests that CPW was more likely selected over other peptides for interaction stability than helicity, assuming the accuracy of these algorithms, which were trained on more canonical CC sequences than those of CPW, CIS1 and CIS21.

In addition to cJun binding prediction, prediction of propensity to form β-fibril aggregates of the top three peptides in CIS display selections R3 and R4, CPW and comparison peptides CIS1 and CIS21 (Table 4.1), was performed using PASTA 2.0 and Zyggregator algorithms (Trovato *et al.*, 2006, Pawar *et al.*, 2005). These quantitative predictions used aggregation propensity as a proxy for preference of peptides for non-specific interactions, to complement qualitative conclusions of non-specificity from hydrophobicity based on amino acid compositions, and was based on the rationale that β-structure (rather than, for example, helical structure) predominates in aggregates of diverse peptides and proteins. These analyses suggested that low stringency selection peptides were generally less aggregation prone (non-specific) than high stringency peptides, particularly based on PASTA 2.0 scores. Further, CPW is expected to be less aggregation prone than the majority of the top three peptides from CIS display including CIS1, and than CIS21, whilst being slightly more prone than control peptides cJun, cFos and FosW which are known not to aggregate at high concentrations. The lesser aggregation propensity/non-specificity of CPW could partly explain its preferential selection during PCA over other peptides enriched by CIS display. It also suggests that CIS display had enriched more aggregation prone peptides which may be non-specific binders of cJun (see 4.3.1 for further discussion).

Based on predicted interaction favourability, CPW outperforms CIS1, the most enriched peptide after CIS display low stringency selection that was

outcompeted by CPW in the subsequent PCA. Thus, predictions by these algorithms support CIS→PCA as successfully selecting CPW as a high affinity cJun binder. Because CIS1 was also the winner of CIS→PCA from the high stringency CIS selection, suggested to have been the less efficient selection, its poor predicted interaction properties relative to CPW again support the success of the low stringency selection.

4.2.5.4 Sequence alignment evidence of CC formation

Sequence alignments were performed to assess the similarity of CPW, CIS21 and CIS1 to CC domains of cFos and FosW, compared to Fos homologues that align well to both cFos and FosW. Manual alignments were performed assuming that all peptides bound to cJun with C-termini aligned, as anticipated in designing the 3.5h library and expected for Fos homologues (Mason *et al.*, 2006) (Figure 4.5A and B). The suitability of these alignments was confirmed by BLASTP alignments using a PAM-30 similarity matrix (most suitable matrix for the length of sequences analysed, see also 2.8.8) (Altschul *et al.*, 1990), which identified C-terminal alignments as scoring highest for identity and similarity (Figure 4.5C and D). CPW, CIS21 and CIS1 all align with cFos (CPW 69 %, CIS21 72 %, and CIS1 84 % similarity) and with FosW (CPW 81 %, CIS21 88 %, and CIS1 91 % similarity) with high percentage similarities along their entire lengths, suggesting that CPW, CIS21 and CIS1 represent CC-like sequences. Furthermore, consensus sequences generated from CPW, CIS21 and CIS1 in alignment with cFos and FosW resemble the generic CC protein consensus sequence n-p-p-n-p-p-p where “n” denotes non-polar and “p” denotes polar residues (Mason and Arndt, 2004), with minor deviations only at one ***f*** and one ***c*** position, where hydrophobic residues are sometimes found in natural LZs (Crooks, 2013).

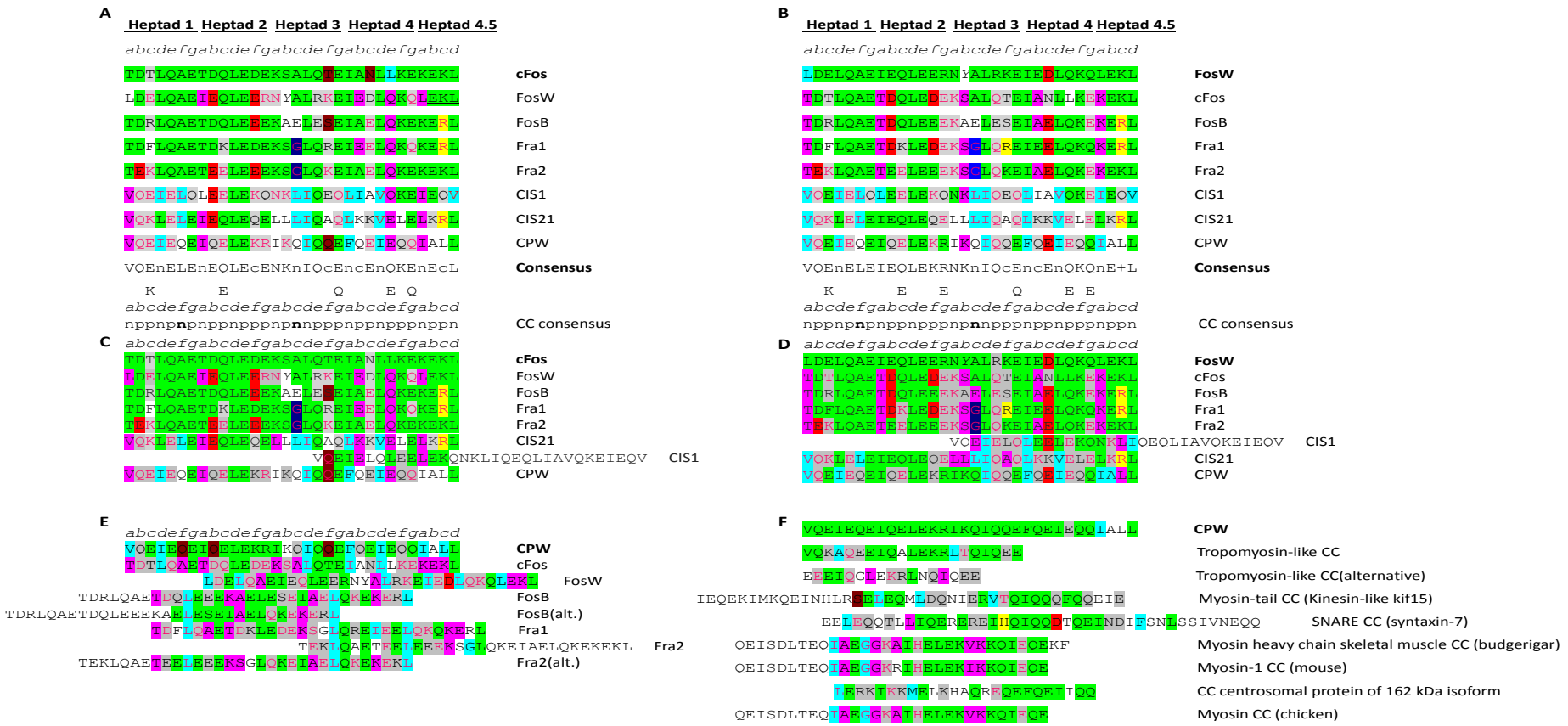


Figure 4.5: Similarity of CPW, CIS21 and CIS1 to known CC proteins (legend continues on next page). Colours for residues in aligned sequences denote the following: green for identity, blue for hydrophobic similarity, yellow for positive charge similarity, red for negative charge similarity, brown for polar (uncharged) similarity, grey for polar/charged similarity, and pink (highlight or text colour) for residues different to that of the query sequence but identical across two or more peptides. Underlined residues were fixed in the library design. The Y in FosW is in italics to denote that it was not a residue selected from a library, and has not been aligned. **Legend continues overleaf.**

Figure 4.5 continued: Similarity of CPW, CIS21 and CIS1 peptides to known CC proteins. Consensus sequences were derived from residue similarity colourings, except fixed residues which were not considered. For consensus sequences, n denotes non-polar residues, p denotes polar residues, + denotes positively charged residues, c denotes charged/polar residues, and X denotes non-polar/polar (any) residues. For BLASTP alignments, the longest stretch of similar amino acids with no gaps are displayed. **A** and **B** Manual sequence alignments for CPW, CIS1 and CIS21 peptides against query sequences cFos or FosW, with Fos homologues (and FosW for alignment to cFos) as reference alignments, based on C-terminal alignment of each peptide with cFos or FosW. Consensus sequences based solely on similarity between peptides CPW, CIS1, CIS21 and cFos or FosW have been constructed, and are compared to the CC consensus n-p-p-n-p-p-p (Mason and Arndt, 2004), with positions not conforming to this pattern in bold text. **C** BLASTP sequence alignments of peptides vs. cFos using standard settings (see 2.8.8 for details) (Altschul *et al.*, 1990). **D** BLASTP sequence alignments of peptides vs. FosW using standard settings. Where BLASTP offered alternative alignments to the C-terminal alignments in **A** and **B**, these have fewer positions of similarity with cFos or FosW than C-terminal alignment, such that C-terminal alignments are preferred. **E** BLASTP alignment of all Fos homologues and FosW to CPW to corroborate C-terminal alignments, with alignments shifted by an entire heptad having lower similarity. **F** Example BLASTP alignments of CPW to CC proteins retrieved from the UniProtKB database. CC protein Accession.Version number identifiers are as follows: PREDICTED: tropomyosin-like [Tetranychus urticae] XP_015795318.1; kinesin-like protein kif15 [Stylonychia lemnae] CDW74084.1; syntaxin 7 [Candida albicans P60002] KHC49515.1; PREDICTED: myosin heavy chain, skeletal muscle, adult-like, partial [Melopsittacus undulatus] XP_005141824.1; myosin-1 [Mus musculus] NP_109604.1; PREDICTED: centrosomal protein of 162 kDa isoform X2 [Callorhinchus milii] XP_007906621.1; myosin, partial [Gallus gallus] CAA23712.1.

CPW was also aligned to all Fos homologues and FosW using BLASTP (Figure 4.5E). CPW aligns with high similarity (69 – 81 %) along its full length to Fos homologues, and all homologues align with heptads in-register. In the cases of FosB and Fra2, alignments are shifted by units of whole heptads, though these alternative alignments have lower similarity than the manual C-terminal alignments in Figure 4.5A and B, again supporting the latter as the most optimal. Finally, BLASTP was used to search for regions of similarity shared between CPW and protein sequences within the UniProtKB database. 104 proteins generated significant alignments based on the search criteria (see 2.8.8), again using the PAM-30 similarity matrix, of which alignments were identified with known CC proteins such as those containing tropomyosin-like, myosin-like, and SNARE-like domains (Figure 4.5F). These findings again support CPW as capable of forming CCs.

4.2.6 *In vitro* characterisation of CPW–cJun interaction

CPW demonstrated high affinity binding to cJun *in vitro* to form a CC, thus demonstrating the success of CIS→PCA selection. CD spectroscopy scans (Figure 4.6) were used to assess α-helicity and CC formation, CD thermal denaturations (Figure 4.6 and Table 4.3) were used to assess stability of formed complexes, and ITC (Figure 4.6 and Table 4.3) was used to more accurately evaluate binding free energy and decompose binding into enthalpic and entropic contributions, as described in 3.2.2. CPW is compared against FosW, $FosW_{CANDI}$, 4hFosW and native cFos–cJun AP-1 where CD and ITC data was available (Mason *et al.*, 2006, Mason *et al.*, 2007b, Crooks *et al.*, 2011, Worrall and Mason, 2011).

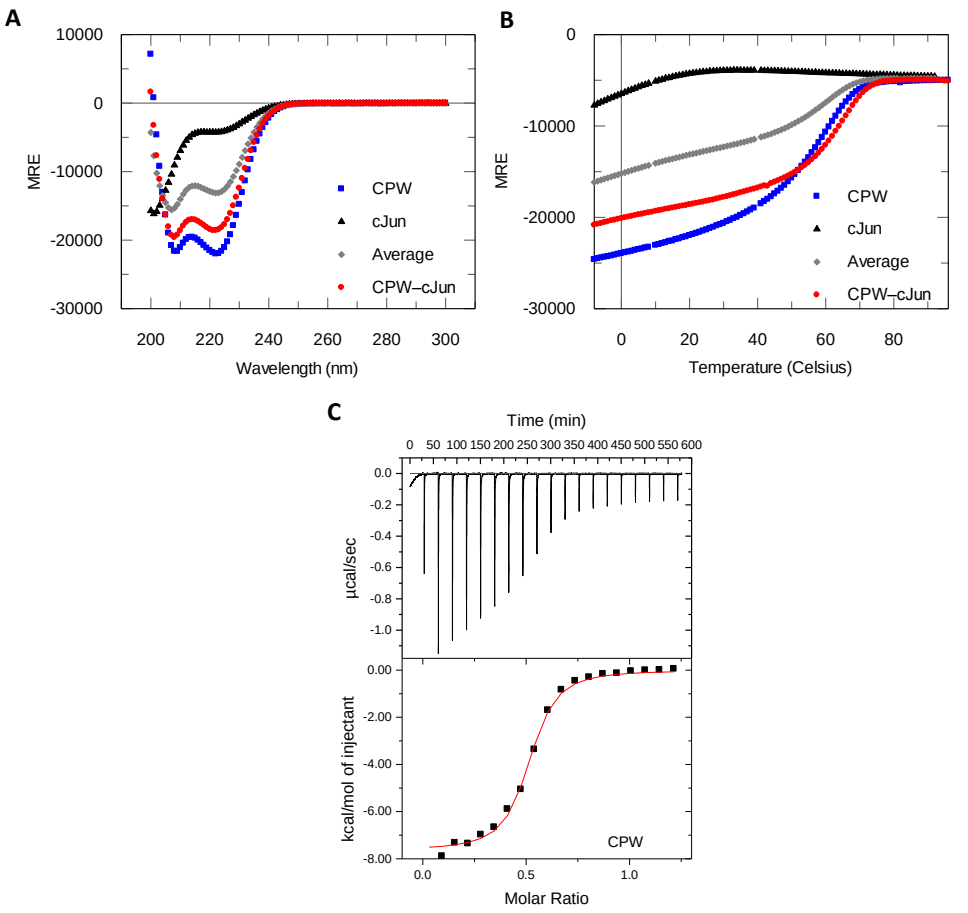


Figure 4.6: CPW binds cJun with high affinity. Biophysical characterisation of the CPW–cJun interaction by circular dichroism spectroscopy (CD) scans (**A**) and thermal denaturation (**B**), and isothermal titration calorimetry (ITC) (**C**). **A** Data reported as change in mean residue ellipticity (MRE; units deg cm^2 $dmol^{-1}$, to allow for comparison between peptides of different lengths), as a function of CD ellipticity over the wavelength range 200 – 300 nm. Blue squares: peptide alone, red circles: equimolar mix of peptide with cJun, grey diamonds: average of cJun and peptide alone, black triangles: cJun alone. Peptides interact with cJun where MREs for peptide:cJun mixes exceed that of the calculated average trace of peptide and cJun only traces (which represents non-interaction). MRE was calculated using Equation 3 (see 2.9.6.1). cJun traces from Mason *et al.* (2006) are in excellent agreement with that reported here. **B** Data reported as change in MRE as a function of CD ellipticity at 222 nm with temperature. Colours are as for CD scans. Peptides interact with cJun where MRE and/or melting temperature (T_m) for peptide:cJun mixes exceed that of the calculated average trace of peptide and cJun only traces (which represents non-interaction). Heteromeric and homomeric traces were fit to Equations 6 – 13 (see 2.9.6.2) to produce data in Table 4.3. **C** Isothermal titration calorimetry raw isotherms (top panels) and fitted data (bottom panels) (both baseline corrected). On the fitted data plot, the solid line represents the fit generated by non-linear least-squares fitting to Equation 14 (see 2.9.7).

Table 4.3: CPW displays high affinity binding to cJun, similar to that of previous cJun antagonists and substantially higher than native cFos. Thermodynamic parameters for CPW in interaction with cJun compared to cJun antagonists FosW (CD values from Worrall and Mason, 2011), 4hFosW (values from Crooks *et al.* 2011), and $FosW_{CANDI}$ (values from Mason *et al.* 2007b), and native cFos (values from Worrall and Mason, 2011). CPW is also compared to less successful CIS→PCA high stringency winner CIS1, and CIS21, the 21st most frequent peptide in the CIS low stringency selection. Data from CD and ITC measurements. θ is raw CD ellipticity (mdeg). Fractional helicity was calculated using Equation 4 (see 2.9.6.1). TΔS is calculated as ΔH – ΔG from the Gibbs-Helmholtz equation (Equation 11, see 2.9.6.2). CD values are from representative single measurements, typically reproducible in biological replicates to ± 1 °C for T_m, within 5 % for fractional helicity and 222:208 ratio ($\theta_{222/208}$), and within 10 % for K_d and ΔG (data not shown). ITC values are the arithmetic mean of two to three independent titrations ± SDs, except values for 4hFosW–cJun, $FosW_{CANDI}$–cJun and cFos–cJun (single titrations and fitting errors). *'low salt' buffer (10 mM KH_2PO_4/K_2HPO_4, no KF), other peptides assayed in KPP buffer ("low salt"+ 100 mM KF). "ND" = not determined. Values given to 2 s.f..

Complex	T_m (°C) (CD)	Peptide–cJun Fractional Helicity (%) (CD)	$K_{d(20°C)}$ (nM) (CD)	$\Delta G_{(20°C)}$ (kcal/mol) (CD)	$K_{d(20°C)}$ (nM) (ITC)	N interaction stoichiometry (ITC)	$\Delta G_{(20°C)}$ (kcal/mol) (ITC)	$\Delta H_{(20°C)}$ (kcal/mol) (ITC)	$T\Delta S_{(20°C)}$ (kcal/mol) (ITC)	Peptide–cJun $\Theta_{222/208}$ (CD)
cFos–cJun	16	28	320,000	-5.5	27,000	1.1 ± 0.01	-6.1 ± 0.39	-0.82 ± 0.36	5.32± 0.53	0.75
FosW–cJun	63	37	4.0	-11	39 ± 11	0.99 ± 0.08	-9.9 ± 0.16	-10 ± 0.42	-0.46 ± 0.46	1.0
4hFosW–cJun	49	60	490	-8.5	480	1.1 ± 0.01	-8.8 ± 0.10	-14 ± 0.20	-5.3 ± 0.20	0.99
$FosW_{CANDI}$–cJun	52	54	ND	ND	ND	ND	ND	ND	ND	1.0
CPW–cJun*	63	53	0.27	-13	750 ± 270	0.48 ± 0.03	-8.2 ± 0.21	-7.4 ± 0.67	0.81 ± 0.81	0.95
CIS1–cJun*	34	31	14,000	-6.5	25,000 ± 13,000	0.11 ± 0.02	-6.2 ± 0.29	-15 ± 4.42	-9.2 ± 4.6	0.71
CIS21–cJun*	68	48	0.73	-12	770 ± 230	0.52 ± 0.04	-8.2 ± 0.18	-3.7 ± 0.21	4.6 ± 0.37	0.92

The heteromeric trace of CPW with cJun displayed a helical signature with minima at 222 and 208 nm (Figure 4.6), and a 222:208 ratio indicative of CPW–cJun CCs (Table 4.3). This trace far exceeds the non-interaction calculated average of homomeric CPW (slightly more helical than CPW–cJun, see Appendix Table 4.3) and cJun, evidencing CPW–cJun interaction. CPW–cJun fractional helicity is increased by +16 % vs. FosW–cJun, is similar to $FosW_{CANDI}$–cJun and 4hFosW–cJun, and is increased by +25 % vs. native cFos–cJun. Thus, CD scans support a CPW–cJun CC interaction.

CPW heteromeric (and homomeric) CCs displayed cooperative thermal denaturation (Figure 4.6) that fit extremely well to a 2-state unfolding model for CCs ((Mason *et al.*, 2007a) and references therein). Highly helical CPW–cJun CCs were very thermally stable, with a T_m of 63 °C, and had a higher T_m than homotypic interaction (Figure 4.6). The CPW–cJun thermal denaturation trace, as for CD scans, far exceeds the non-interaction average with regard to helicity and T_m. The thermal stability of CPW–cJun compares very favourably with CCs of cJun antagonists previously derived, being identical to that of FosW–cJun, and higher than 4hFosW–cJun and $FosW_{CANDI}$–cJun by 14 °C and 11 °C respectively. Compared to the native cFos–cJun AP-1 complex, CPW–cJun is more stable by a ΔT_m of 47 °C, indicating that *in cellulo* CPW would outcompete cFos (or indeed any other Fos or Jun homologue) to sequester cJun into more stable complexes, thus reducing cFos–cJun formation.

ITC elicited the expected sigmoidal binding curve of a high affinity CPW–cJun interaction (Figure 4.6), whose fitting derived a K_d of 750 ± 270 nM, which translates to a ΔG of -8.2 ± 0.21 kcal/mol. Enthalpic contributions to ΔG dominated binding (-7.4 ± 0.67 kcal/mol) relative to the entropic term (0.81 ± 0.81 kcal/mol). These values

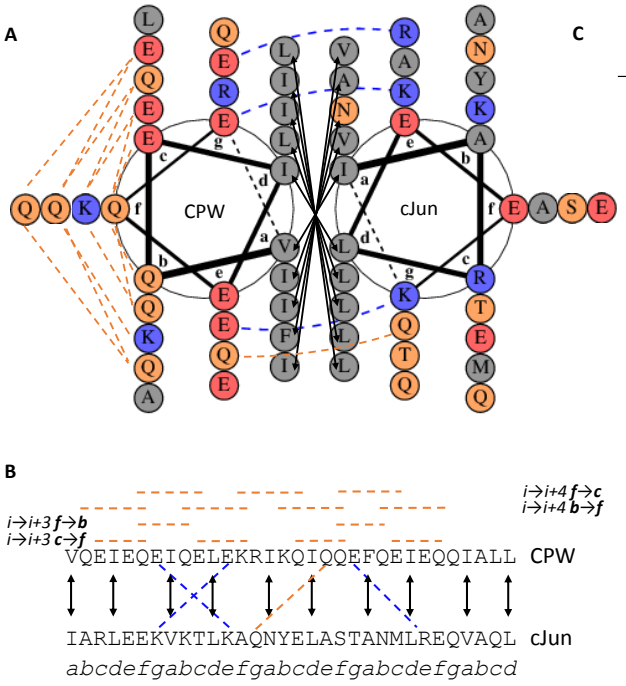
comfortably fit in the range for CC interactions including FosW, 4hFosW and cFos, and the enthalpically-driven CPW–cJun interaction is typical of CCs (Worrall and Mason, 2011, Crooks *et al.*, 2011). It is noted that there is a discrepancy between affinities and free energies for CPW–cJun measured by CD and ITC. This could be attributable to the inaccuracy of CD determination of binding affinity relative to ITC, as previously observed for Fos and Jun antagonists (Worrall and Mason, 2011), and/or an alternative oligomeric state than the expected dimer (Privalov, 1979) (see 4.2.8). These suggestions will require further investigation. Despite this, CD and ITC agree on a high affinity interaction between CPW and cJun. The ITC-derived ΔG of CPW for cJun interaction compares well with that of FosW and is similar to that of 4hFosW, being 35-fold higher than the affinity of cFos for cJun. This supports the suggestion that *in cellulo* CPW would be an effective cFos–cJun antagonist.

Additionally, CPW displayed higher helicity (+22 %), a much higher T_m (+29 °C), and consequently a 34-fold lower K_d by ITC than CIS1, supporting CIS→PCA low stringency selection of CPW over CIS1. Finally, CPW performed as well as peptide CIS21 (Table 4.3), whilst being more soluble *in vitro*: CIS21 was assayed by CD at 20 μM and by ITC at 70 μM, and data was normalised for concentration using Equation 3 in 2.9.6 or using Equation 14 in 2.9.7. The greater solubility of CPW was also suggested by aggregation analysis (see 4.2.5.3), and may have favoured selection of this peptide *in cellulo*, amongst other properties, over CIS21.

4.2.7 CPW–cJun interaction; selected amino acids

Residue selections in CPW, inferred interhelical interactions with cJun residues, and intrahelical interactions potentially stabilising the CPW helix, assuming

formation of a parallel dimeric CC, are presented in Figure 4.7. It is noted that because the diversity of the 3.5h library was not fully covered, some residue choices may have been limited to what was sampled from the library at the start of selection.



C

Proposed interacting positions	Proposed interhelical interaction	ΔΔG relative to A–A (kcal/mol/dimer)
***a*1–*a'*1**	V–I	-6.1 to -6.3
***d*1–*d'*1**	I–L	-5.7 to -5.8
***e*1–*g'*0**	Not possible	-
***g*1–*e'*2**	E–K	-0.9
***a*2–*a'*2**	I–V	-6.1 to -6.3
***d*2–*d'*2**	L–L	-5.2
***e*2–*g'*1**	E–K	-1.4
***g*2–*e'*3**	R 'paired' with A	-0.2
***a*3–*a'*3**	I–N	-0.4 to -0.6
***d*3–*d'*3**	I–L	-5.7 to -5.8
***e*3–*g'*2**	Q–Q	-1.2
***g*3–*e'*4**	E–R	-1.3
***a*4–*a'*4**	F–A	ND
***d*4–*d'*4**	I–L	-5.7 to -5.8
***e*4–*g'*3**	E 'paired' with T	ND
***g*4–*e'*4.5**	Not possible	-
***a*4.5–*a'*4.5**	I–V	-6.1 to -6.3
***d*4.5–*d'*4.5**	L–L	-5.2

Figure 4.7: Possible interactions stabilising the assumed dimeric CC interaction between CPW and cJun. A DrawCoil helical wheel diagram (Grigoryan, 2016). Core interhelical ***a*–*a'*** and ***d*–*d'*** interactions are shown as black solid-line arrows, and expected stabilising interhelical $\boldsymbol{e}_{i}$–$\boldsymbol{g'}_{i-7}$ and $\boldsymbol{g}_{i}$–$\boldsymbol{e'}_{i+7}$ interactions are shown as blue curved dashed lines. A possible interhelical polar-polar interaction (***e*3–*g'*2** Q–Q) is shown as an orange curved dashed line. All possible *i→i+3* and *i→i+4* intrahelical interactions between outerface ***b***, ***c*** and ***f*** residues are displayed with straight orange dashed lines; note that formation of some intrahelical interactions will preclude formation of others, and that intrahelical interactions can form between outerface residues and ***e*** and ***g*** residues (not drawn) (O'Shea *et al.*, 1991). Residues are coloured grey for hydrophobic, red for positively charged, orange for polar, and blue for negatively charged. **B** Linear sequence view of helical wheel diagram in **A** using same colouring convention for proposed interactions, except interhelical ***e*–*g'*** interactions are shown as straight rather than curved lines. Heptad register shown in italics. **C** Contributions to interaction ΔG from selected residues in CPW assuming interaction with cJun residues, with Ala–Ala as the reference state, mimicking no selection. ΔΔG energies for ***a*–*a'*** and ***d*–*d'*** interactions from Acharya *et al.* (2006), and for ***e*–*g'*** and ***g*–*e'*** interactions from Krylov *et al.* (1998), with two values depending on residue orientation. ND = not determined.

The majority of selected core residues are those that contribute most to dimerisation ΔG from the library options provided, where reference data is available (Acharya *et al.*, 2006). These are mostly Ile, selected at three out of five ***a*** positions and three out of five ***d*** positions. At ***a*** positions, Ile side chains could favourably adopt parallel packing orientations typical of these positions in a dimer; however, perpendicular packing orientations typical in dimers at ***d*** positions are far less favourable for Ile (Harbury *et al.*, 1993, Harbury *et al.*, 1994), making Ile an interesting choice. Three ***a*** (including an Ile) and two ***d*** position choices are not the most optimal selections. The first is ***a***1 V, proposed to form an ***a***1–***a'***1 V–I interaction, which contributes only slightly less to dimerisation ΔG than I–I, and Val is a Cβ-branched residue for favourable parallel packing at this position. The inferred ***a***3–***a'***3 interaction of I–N is the fourth best of all potential interactions previously studied for dimerisation ΔG below N–N (approx. 2 kcal/mol/dimer more stable), N–A, and L–N, and makes a very small favourable contribution to ΔG. Finally, at the CPW ***a***4 position, Phe has been chosen opposite cJun's ***a'***4 A. This selection cannot be scored due to insufficient data from double mutant thermodynamic cycles. There is potential for van der Waals' interactions between the Phe and Ala, but the contribution of these to stability is unclear, as is to what extent the large Phe side chain can be accommodated in a dimeric core. Finally, ***d***2–***d'***2 and ***d***4.5–***d'***4.5 L–L contribute slightly less to ΔG than I–L but Leu should adopt favourable perpendicular packing arrangements to encourage dimerisation (Harbury *et al.*, 1993, Harbury *et al.*, 1994).

Three out of five ***e*** and ***g*** positions that would be expected to form interhelical $\boldsymbol{e}_{i}$–$\boldsymbol{g'}_{i-7}$ and $\boldsymbol{g}_{i}$–$\boldsymbol{e'}_{i+7}$ interactions appear to do so (Figure 4.7). Residues selected at ***g***1, ***e***2, and ***g***3 would be expected to make interhelical contact, and ***e***3 Q could perhaps

make contact with the cJun ***g'***2 Q, with these four residues being the most optimal for dimerisation free energy (Krylov *et al.*, 1998). At the remaining ***e***4 position that could form an interhelical interaction, an E residue is 'paired' (interaction unlikely) with T, a combination that has not previously been scored. Selection of R at ***g***2 also makes a small contribution to ΔG when 'paired' with cJun ***e'***3 Ala relative to E, K and Q alternatives. This Arg could participate in a stabilising ***g–a'*** interaction with the ***a'***3 N of cJun, as in cFos–cJun to improve dimer stability (Glover and Harrison, 1995). Alternatively, R may interact with solvent/be attracted to the CPW ***g***3, as observed for the same positions of FosW in the FosW–cJun crystal structure (Chapter 3).

Possible intrahelical interactions between ***b***, ***c*** and ***f*** residues that could potentially contribute favourably to dimerisation ΔG through CPW stabilisation in a helical conformation are depicted in Figure 4.7. The expected preference for E, K, Q, R and A rather than I or L was observed, with predominance of Q (seven positions) over E (three positions), K (two positions) and A (one position). The ***b***4.5 Ala residue could increase helicity in this poorly helical region of CPW (Figure 4.1) (O'Neil and Degrado, 1990), whilst reasons for other selections will require further investigation.

4.2.8 Oligomeric state: potential trimeric interaction between CPW and cJun

Size exclusion chromatography (SEC) (Figure 4.8) and native polyacrylamide gel electrophoresis (native PAGE) (Figure 4.9) were performed to assess the oligomeric state of CPW–cJun complexes, and confirm CD and ITC analysis and residue selection discussion based on a dimeric CC. SEC data, however, suggested possibility of a trimeric complex, whilst native PAGE data confirmed CPW–cJun interaction but was inconclusive of oligomeric state.

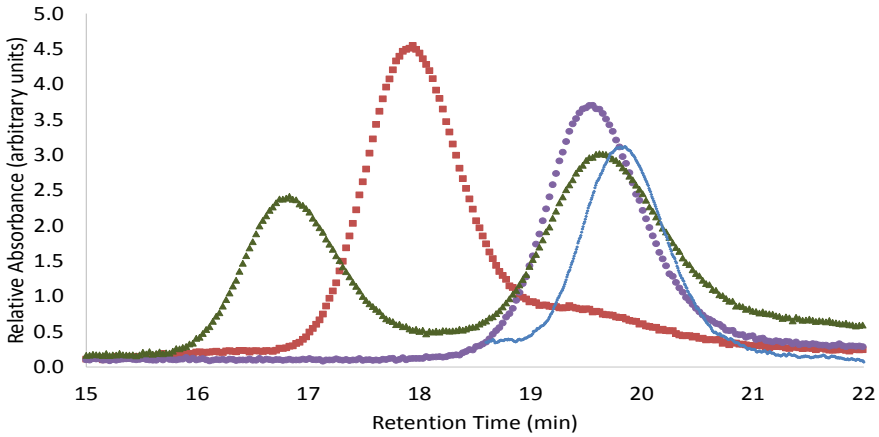


Figure 4.8: Possible CPW heterotrimer formation with cJun using size exclusion chromatography. Relative absorbance at 280 nm vs. retention time (min) for dimeric control FosW:cJun 1:1 stoichiometric mixture (red squares), monomeric control cFos (blue diamonds), cJun (purple circles) and CPW:cJun 1:1 stoichiometric mixture (green triangles) samples, loaded at 150 µM total peptide concentration (same concentration as CD analysis).

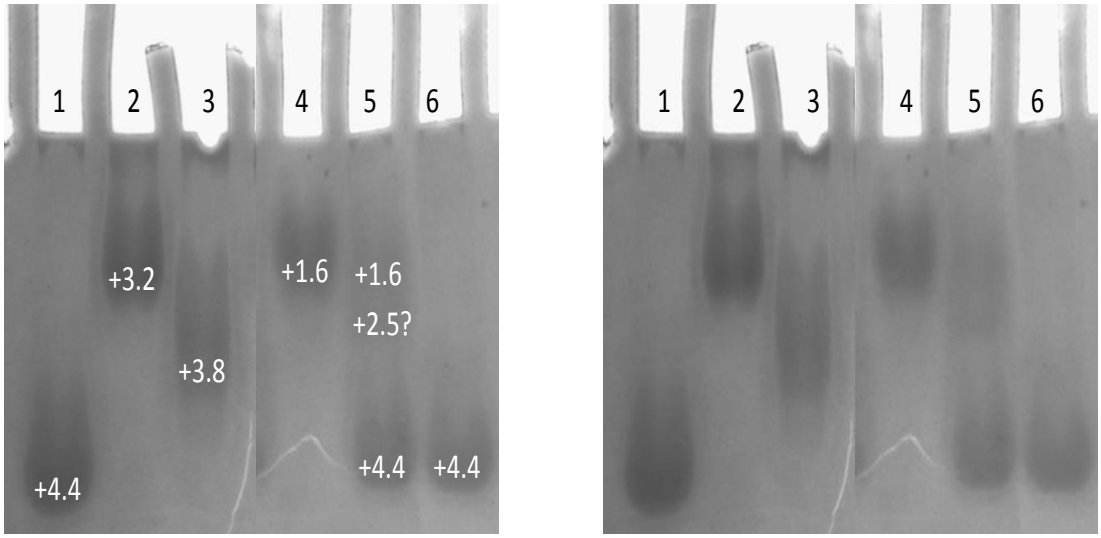


Figure 4.9: CPW interaction with cJun supported by native polyacrylamide gel electrophoresis. Samples were loaded at 600 µM. Lane 1: cJun, Lane 2: FosW, Lane 3: FosW–cJun (heterodimeric CC control), Lane 4: CPW, Lane 5: CPW:cJun 1:1 stoichiometric mix, Lane 6: cJun (300 µM). Expected charges for peptides alone/homodimers (assuming same charge on either), or heterodimers (average of peptide alone and cJun charges) are laid over the presumed corresponding bands in white text. Note the question mark for the CPW–cJun band between +1.6 (CPW) and +4.4 (cJun); the distance this band has migrated relative to +1.6 and +4.4 bands could support a charge of +2.5 for a dimer or +3 for an CPW–cJun trimer. Un-annotated gel shown on the right for clarity of band viewing.

SEC was performed on an equimolar mixture of CPW and cJun at the same concentration as CD analysis (150 μM total peptide concentration), and gave two A_{280} absorbance peaks at retention times of 16.8 min ("Peak 1") and 19.6 min ("Peak 2"). An equimolar mixture of FosW with cJun, used as a heterodimeric control, gave a single peak at a retention time of 17.9 min. cJun alone gave a single major peak at 19.6 min suggested to represent predominantly cJun monomers based on a similar retention time to monomer control cFos (Smeal *et al.*, 1989) at 19.9 min. The earlier retention time of Peak 1 relative to the major peak of FosW with cJun suggests a higher order oligomer is formed between CPW and cJun than between FosW and cJun, considering that CPW–cJun and FosW–cJun dimers should be of such similar MWs (ΔMW 24 Da) as to be unresolvable by SEC. Peak 1 must represent a complex that includes cJun molecules, as CPW alone has no chromophores absorbing significantly at 280 nm, and so is unable to generate an absorbance trace at this wavelength. The shift in retention time for cJun molecules contributing to Peak 1 at 16.8 min from the major peak of cJun alone at 19.6 min supports a heteromeric complex formed through interaction with CPW. Extrapolation of FosW–cJun and cFos control MWs to the retention of Peak 1 for CPW with cJun suggests a MW of ≈11,000 Da. This is in better agreement (≈+1900 or +1600 Da) with expected MWs for possible trimeric species CPW–CPW–cJun or CPW–cJun–cJun than a CPW–cJun dimer. Peak 2 at 19.6 min could therefore represent leftover cJun not complexed with CPW.

Further, Logicoil *in silico* analysis (Vincent *et al.*, 2013) predicts that CPW prefers homotrimerisation over parallel or antiparallel dimerisation, or tetramerisation (Appendix Table 4.4). Logicoil is only able to predict oligomeric state

preference for homomeric complexes, however this prediction does suggest that CPW residues do not strongly encourage dimer formation.

The suggested trimeric state of CPW–cJun is, however, far from conclusive. PCA is expected to favour selection of parallel dimers due to the nature of refolding of the two mDHFR fragments fused to library peptide and target (CIS display, on the other hand, does not apply any preference for binding orientation). Thermal denaturation of CPW–cJun (Figure 4.6) showed a dissociation with apparent 2-state cooperativity, typical of dimers ((Mason *et al.*, 2007a) and references therein). However, trimers may also undergo 2-state unfolding as an alternative to 3-state unfolding (Dragan *et al.*, 2004). ITC for CPW–cJun (Figure 4.6) similarly produced a monophasic transition that fit well to a single site binding model typical of dimers and not to two site binding models for 3-state unfolding (Appendix Figure 4.5) (Wiseman *et al.*, 1989), though again 2-state trimer unfolding is not precluded. The helical wheel diagram discussed above for a CPW–cJun dimer (Figure 4.7) is favourable based on core and electrostatic interhelical interactions; however, a helical wheel diagram for possible heterotrimers (Appendix Figure 4.6) also has some features supporting trimer formation. Further investigation as to the oligomeric state of CPW–cJun complexes is required to confirm these suggestions.

4.3 Discussion

4.3.1 Success of CIS→PCA as a novel library screening strategy

CIS display and PCA *in vitro* and *in cellulo* library screening approaches have been combined for the first time to assess whether together they could constitute an effective system for identification of novel AP-1 antagonists with attractive

intracellular stability from huge peptide libraries. It was expected that CIS→PCA could benefit from considerable overlap in the capabilities and shortcomings of isolated CIS and PCA systems if CIS display and PCA systems were compatible.

The greatest concern for display system combination is that it is possible that selection pressures applied to peptides to select them for binding to a target will differ in the different environments of *in vitro* and *in cellulo* systems. More so in *in vitro* systems than in *in cellulo* systems, peptides can experience an additional secondary selection pressure besides the primary selection pressure for specific binding to the target. This secondary selection pressure selects peptides that non-specifically interact/aggregate with components of the selection and which are difficult to remove by washing. This could potentially mean that CIS display would enrich non-specific (highly hydrophobic)/aggregating peptides, which would not be expected to be selected in PCA, which selects against non-specific peptides as these do not afford host cell growth through sufficient specific interaction with the target and mDHFR recombination (Mason *et al.*, 2006). In the worst case scenario, selection pressures in CIS display and PCA could therefore antagonise each other.

The strength ("stringency") of this secondary undesirable selection pressure in CIS display, relative to the primary selection pressure, depends on the stringency of the latter compared to the affinity of peptides in the library for the target. In the case of CIS→PCA, primary selection stringency was dependent on the availability of target for interaction, partly determined by the quantity of cJun coated on magnetic beads. In addition to adequate saturation ("blocking") of *in vitro* selection components that peptides could bind to using blocking agents such as BSA, more

cJun target in the CIS selection can increase the chance for peptide–target interaction and thus decrease the chance for non-specific binding (secondary selection pressure). However, this also results in a lower primary selection stringency, such that a larger pool of binders with affinities ranging to higher K_d values is selected (the selection is less efficient). A higher primary selection stringency (less cJun target) conversely selects for peptides of decreased K_d values, but may suffer from stronger secondary selection of non-specific binders, particularly if the quantity of target is much less than the K_d of the highest affinity library peptide. Thus, two cJun quantities designated “low stringency” and “high stringency” were applied in CIS display, guided by previous CIS selection of FosW (Rao, T., 2010, unpublished results). This covered a range of primary stringencies, with the anticipation that one stringency would represent a sufficient balance between specific and non-specific selection to allow PCA to then select from this peptide subset a high affinity, specific binder of cJun.

Of the two stringencies, the low stringency CIS selection appears to have been more appropriate for specific binder selection than the high stringency condition, though even the low stringency condition was not optimal. This stringency corresponded to a cJun concentration of 290 nM (noting that local cJun concentration on beads is much higher than this), such that peptides with K_d values ≤290 nM would be expected to be selected, whilst the high stringency condition corresponded to 70 nM cJun, which should have selected peptides with K_d values ≤70 nM. The low stringency condition should therefore have selected peptides with K_d values lower than those measured for the top most enriched peptide CIS1 (ITC K_d 25 ± 13 µM), peptide CPW at position 71 (ITC K_d 750 ± 270 nM) and peptide CIS21 at position 21 (ITC K_d of 770 ± 230 nM). This suggests that the primary selection

stringency was probably above the affinity threshold of peptides in the 3.5h library, such that secondary selection pressure for non-specific binders was also high. The preferential CIS display enrichment of CIS1 over CPW at this stringency, and the lower cJun-binding affinity and higher aggregation propensity of CIS1 indicative of non-specific binding, supports this suggestion. However, peptide CPW was still enriched sufficiently (to position 71) to ensure transferral to PCA, where it could go on to be selected as the assay winner, and display *in vitro* high affinity binding with a K_d relatively close to the maximum possible.

Conversely, the primary stringency of the high stringency selection was likely so far above the affinity threshold of peptides in the 3.5h library that secondary selection pressure for non-specific binders predominated. This is supported by aggregation propensity analysis suggesting that this CIS display condition may have selected highly hydrophobic, more aggregation-prone sequences than the low stringency selection at the top three positions, and that CIS1, again the most enriched peptide, did not display a K_d ≤70 nM or lower than CPW, contrary to what would be expected if the primary selection stringency had predominated. Furthermore, the fact that CIS1 was selected as the winner in the following PCA indicates that, unlike the low stringency selection, no peptides enriched by CIS display and transferred to PCA could outcompete this peptide, suggesting that the selected peptides of the high stringency selection were generally not specific, high affinity binders. The low stringency condition was therefore the most appropriate for the affinity of 3.5h library peptides, but could be optimised in the future.

Before this study, it was previously unknown whether selection conditions for peptides against a target were different enough in *in vitro* and *in vivo* systems to preclude their successful combination. This study demonstrates for the first time effective combination of an *in vitro* and *in cellulo* library selection system, without such antagonism of selection pressures as to prevent selection of high affinity cJun binder CPW from a diverse library of ≈6.1 x 10^{10} peptides, one of the largest libraries screened against cJun to date.

The 3.5h library was designed to identify an alternative template to FosW for further peptide development efforts, and probe the favourability and function of aromatic residues at core positions, and ***b***, ***c*** and ***f*** position residues. CPW, with its sufficiently different sequence compared to FosW represents such an alternative template, whilst selection of a Phe residue at CPW's ***a***4 position and selections at ***b***, ***c*** and ***f*** residues will be explored with future structural investigation and confirmation of oligomeric state. These insights will contribute to our growing knowledge of the structural and sequence determinants of CC formation, which can be exploited in the design of new libraries for selection of modulatory peptides of disease-related CCs. Finally, the CPW peptide is potentially interesting for further development towards therapeutic, diagnostic or synthetic biology applications.

CIS→PCA has attractive prospects with future refinement. Library coverage of CIS display is amongst the highest of all screening systems described to date (Ullman *et al.*, 2011), and ability to screen *in cellulo* as well is expected to generate hits with better therapeutic properties than *in vitro* selection alone. CIS→PCA may in the future prove to be an attractive alternative to powerful but complex *in vivo* phage

display, where library peptides are selected for affinity and desirable drug-like properties through administration to animal disease models (Arap *et al.*, 2002).

4.3.2 3.5h library construction issues

Construction of the 3.5h library resulted in approx. 17 % of the peptides screened by CIS display being longer than designed. It was hypothesized that longer peptides could be preferentially selected over correct length peptides through formation of more points of enthalpic interaction with the target. The selection of CPW, a 4.5-heptad peptide over 2.4 x 10^{10} correct length peptides presented to cJun in CIS display supports this, and corroborates the conclusion in Chapter 3 that affinity is spread along the length of the cJun interface, making identification of shorter peptides of high affinity challenging.

Following identification of this length problem in the 3.5h library QC deep sequencing, steps were taken to select correct length peptides. A new library without ProxiMAX length issues, "3.4hFosW", was constructed by PCR using a degenerate library primer. This was screened by CIS→PCA in parallel to the 3.5h library, and the results from this screening are explored in Chapter 5. Additionally, after CIS display of the 3.5h library, a second PCA was setup in parallel to the one described here. In this PCA, all colonies on the Single Step selection plate were picked and coding sequences of selected peptides were sequenced. Only colonies harbouring peptides of the correct length were transferred to Competition Selection in liquid medium. Selective transfer of these colonies to Competition Selection would prevent out-competition by colonies harbouring longer peptides, to identify correct length binders of cJun. The results of this PCA are also presented in Chapter 5.

Though the 3.5h library length problem did forego probing whether shorter cJun antagonists could be selected at least in the initial CIS→PCA screen presented in this Chapter, nevertheless this Chapter demonstrated proof-of-principle of the CIS→PCA technique to desirably identify a novel peptide quite different in sequence to FosW and of high affinity for cJun.

4.3.3 CPW, a novel AP-1 antagonist

CPW has favourable binding characteristics when compared to previously described cJun antagonists FosW, 4hFosW, Jun-d1, C2ds, and A-Fos (Potapov *et al.*, 2015, Pazos *et al.*, 2015, Olive *et al.*, 1997). As described in 4.2.6, the affinity of CPW for cJun by ITC is comparable to FosW and 4hFosW, high affinity antagonists of similar size to CPW. The affinity of CPW is lower than that of computationally designed Jun-d1 and peptide-DNA conjugate C2ds (6 nM and 362 nM as determined by solution FRET and fluorescence anisotropy respectively), and A-Fos, an LZ CC peptide with an acidic extension (0.03 nM as estimated by CD thermal denaturation), assuming comparisons can be made between affinities derived by different biophysical techniques. However, the advantage of CPW is its smaller size; at 37 residues including helix-capping motifs, CPW is shorter than Jun-d1 by five residues, similar in length to 35mer C2ds but lacking 18 bp of dsDNA conjugated to C2ds, and shorter than A-Fos by 26 residues. As such, CPW's affinity compares well with these molecules when antagonist length is taken into consideration, whilst being of a smaller size that may be more attractive for therapeutic use. Finally, compared to the 10 µM *in cellulo* IC_{50} of small molecule DNA-binding inhibitor T-5224 (Aikawa *et al.*, 2008), CPW's ≈750 nM K_d is promising for an *in cellulo* IC_{50} similar to that of T-5224.

Further, by targeting the CC domain of cJun rather than the DNA-binding domain – which is more similar to other peptides in other bZIP transcription factor families than the CC domain (Fujii *et al.*, 2000, Vinson *et al.*, 2002) – CPW could be more specific to cJun-containing oncogenic AP-1 compositions than T-5224.

It is suggested that CPW could potentially bind to cJun to form a trimeric CC, though this requires further investigation, and was unexpected as PCA biases selection towards parallel dimers due to mDHFR refolding restraints. However, it is possible that trimers selected by CIS display and featuring one antiparallel helix could, in PCA, allow the mDHFR fused to the two parallel helices to nevertheless refold in a manner similar to that for a parallel dimer. Indeed only core selection of Asn opposite the Asn of cJun, or Leu opposite the Leu of cJun would likely favour parallel over antiparallel orientation of peptides (Oakley and Kim, 1998, Harbury *et al.*, 1993), with other residue selections conferring little preference. Alternatively, a parallel trimer could increase the local concentration of mDHFR fragments to shift the dynamic equilibrium of mDHFR refolding towards association and thus elicit faster bacterial colony growth, or trimerisation could simply be an artefact of *in vitro* characterisation. Trimerisation has not been observed in screening of previous libraries, presumably as these have fixed Leu at all ***d*** positions and N opposite the N of cJun (Mason *et al.*, 2006, Mason *et al.*, 2007b).

Finally, it is interesting to note that whilst identical sequences to FosW, $FosW_{CANDI}$ and 4hFosW could not have been selected from the 3.5h library, closely related sequences different at an ***f***, ***b*** and ***c*** residue (two ***b*** positions for 4hFosW) could have been selected. Whether these peptides were covered in the CIS→PCA

selections and were outcompeted by CPW and other peptides, or were not covered, would be interesting to investigate further.

4.3.4 Future directions

In the future, library construction by ProxiMAX could be altered to minimise issues leading to library peptides of undesirable lengths. Stringent PAGE purification at several stages during library construction, such as of heptads before their ligation and of heptad ligation PCR products before subsequent use, would help remove longer constructs, though the efficiency of this strategy may not be high. Further, this would significantly reduce yields at crucial stages of the library build, increasing library construction time and expense when constructing highly diverse libraries. Alternatively/in addition, the synthetic cJun target peptide used in CIS display could be altered to favour selection of peptides shorter than cJun itself. Alterations include helical constraints functionalised with solvent-exposed bulky moieties (Smeenk *et al.*, 2012) and placed at the hydrophobic interface of cJun in the N-terminal or C-terminal heptad, to sterically block the corresponding regions of longer peptides attempting to bind there. Alternatively, a cJun target truncated by an N-terminal or C-terminal heptad could be used as the screening target. However, both of these strategies would alter the helicity and supercoiling ability of cJun, with the risk that peptides selected against such targets may not bind to native cJun in a therapeutic context.

Having demonstrated the success of CIS→PCA, future efforts could focus on optimising this technique for maximal benefit. For example, the lower affinity of CPW for cJun relative to that of FosW and 4hFosW could be due to a sub-optimal CIS display selection stringency for peptide affinity. The quantity of cJun target used in

the low stringency selection was similar to that used previously to select FosW from a doped library (Rao, T., 2010, unpublished results). As such, isolation of peptides with FosW-like affinities was expected, assuming they existed in the diversity of the 3.5h library, and were part of the library diversity covered in selections (which would require further investigation). However, the library in the study doped with FosW was also expected to have lower affinity for the target than FosW, such that FosW may have been pulled down quantitatively relative to other peptides (Jalali-Yazdi *et al.*, 2016); using the same quantity of cJun may therefore not have been optimal for selection of the highest affinity peptide from the 3.5h library. More thorough exploration of CIS selection stringency could test the pulldown of previously derived cJun antagonists covering a range of affinities by a wider range (using smaller increments) of cJun target quantities, in order to choose the target quantity capable of pulling down only the highest affinity binders. However, caution must be exercised that primary selection stringency is not too high, as argued has occurred for the high stringency selection in this Chapter particularly. Thus, from the work presented here, a more optimal selection stringency for future CIS→PCA refinement could use <290 nM cJun, or between 290 nM and 70 nM cJun if non-specific binder selection during CIS display could be reduced by other means. Aside from stringency considerations, ProxiMAX codons could also be altered to maximise peptide expression in *E. coli* during the PCA stage to increase the efficiency of CIS→PCA selection.

Future work could investigate minimization of undesirable CIS selection of non-specific binders, particularly if selection stringency is to be increased to select higher affinity binders. To more accurately mimic intracellular conditions experienced by peptides in PCA and by administered therapeutic agents to avoid this

issue, different blocking agents that better remove hydrophobic non-specific peptides from selections could be used, and selections could be supplemented with proteases to remove unstructured peptides (Eldridge *et al.*, 2009). For even more accurate mimicking of in-cell conditions, concentrated cell lysate could be added to CIS display selections. Further work could investigate expedition of CIS→PCA: for example, CIS display selection could be reduced to a single round through deep sequencing of time interval samples to identify peptides with the slowest target-dissociation rates (Jalali-Yazdi *et al.*, 2016).

Finally, further investigation of CPW and the nature of its cJun binding would be informative for future development of this molecule towards an application, design of further libraries for novel AP-1 antagonists, and further rationalisation for CIS→PCA selection of CPW. For example, crosslinking of CPW and cJun, such as via disulfide bonds or thioesters at N- or C-termini, and separation of the CC complexes formed by chromatography would identify the oligomeric state and helix orientation of the CPW–cJun structure (Hadley and Gellman, 2006). CD thermal denaturation could then be performed on pure complexes, to confirm whether the discrepancy between CD- and ITC-derived affinities for CPW–cJun based on equimolar mixing were due to trimer formation. Analytical ultracentrifugation could also determine oligomeric state of CPW–cJun complexes. An X-ray diffraction or NMR structure of the CPW–cJun complex would unequivocally evidence the formation of CCs, the oligomeric state of these complexes, their CPW and cJun composition, and helix orientation. Subsequent elucidation of packing arrangements in the CPW–cJun complex, particularly of the ***a***4 Phe of CPW, could inform future library design regarding inclusion of aromatics at CC ***a*** and ***d*** positions and perhaps derivation of

new CC prediction/design rules. Library designs fixing some ***d*** positions as Leu and providing charged residue options like Arg at ***a*** positions could more strongly encourage parallel dimers for better exploration of the role of ***b***, ***c*** and ***f*** positions, and aromatics at core positions, in these CCs (Harbury *et al.*, 1993, McClain *et al.*, 2002).

Further *in vitro* biophysical analysis of CPW–cJun could include use of fluorescence polarisation for quantification of cJun binding K_d to corroborate ITC measurements of affinity (Kohler and Schepartz, 2001). The specificity of CPW's ability to bind cJun could be assessed by 'dimer exchange CD' using competing Fos and Jun homologues, to evaluate the preference of CPW for cJun (Crooks *et al.*, 2011). Cellular delivery of CPW or transfection into a cell line would enable *in cellulo* AP-1 inhibitory activity to be assessed, either by luciferase assay (Young *et al.*, 1999), or qRT-PCR assessment of AP-1 transcriptional activity as described in Chapter 3.

In conclusion, CIS→PCA combination represents a novel high throughput library screening technique, capable of screening hugely diverse peptide libraries with refinement of desirable therapeutic properties of peptides through *in cellulo* selection, which has identified a novel cJun antagonist. Further investigation and development of this peptide, as well as refinement of CIS→PCA, could generate valuable antagonists for the future therapy of cancers featuring AP-1 dysregulation.

Chapter 5 – Truncated peptides selected from hugely diverse libraries by CIS→PCA exhibit poor binding compared to longer antagonists

5.1 Introduction

Peptides are increasingly proving to be attractive therapeutic agents, with >60 currently in the clinic (Fosgerau and Hoffmann, 2015). Short peptides can demonstrate attractive pharmacokinetics without necessarily conforming to Lipinski's Rule of Fives for small molecule oral bioavailability (Lipinski *et al.*, 1997, Craik *et al.*, 2013), and thus peptide downsizing is a major aim for therapeutic application. Accordingly, therapeutic peptides are typically ≤50 residues in length (Kaspar and Reichert, 2013). Peptidic AP-1 antagonists towards the lower end of this size range could therefore be very attractive for future cancer therapy. However, this poses a significant challenge, not in the least that the AP-1 CC PPI between Jun and Fos proteins involves a surface approximately 1800 $Å^2$ in size (O'Shea *et al.*, 1991), compared to helical peptides which span only ≈5.4 Å in length per turn.

Short peptides must also be able to adopt a stable target-binding structure. The shortest natural α-helices in PPIs (heavily biased by those in globular proteins) are on average four turns in length (≈15 residues) (Jochim and Arora, 2010), and are stabilised by amide-carbonyl hydrogen bonds contributing ≈1 kcal/mol each to α-helix formation ΔG (Makhatadze, 2005). This motif has been shortened to a single turn (four residues) through the use of lactam helix constraints (Shepherd *et al.*, 2005). Unmodified peptides shorter than four helical turns may be unable to overcome entropic barriers to α-helix formation, and therefore are unable to form a CC (Yang and Honig, 1995, Kentsis and Sosnick, 1998, Zimm and Bragg, 1959).

Accordingly, the shortest natural CCs involve α-helices of 3 – 4 turns, though again these are found predominantly within globular proteins and at higher oligomeric states than dimers, both factors which stabilise these CCs (Lupas and Gruber, 2005). Dimeric CCs featuring four helical turns and that are not part of globular proteins can be designed *de novo* through maximising intra- and interhelical interactions on both peptides (Burkhard *et al.*, 2002), but may represent the lower limit for CC length as peptide truncation results in substantial loss of ΔG (De Crescenzo *et al.*, 2003). It is likely that there exists a 'sweet spot' between long peptides which form stable α-helices but display poor druggability, and very short peptides which are more therapeutically attractive but may be unable to form helices and CCs.

Peptide library display and selection systems are an efficient way to identify peptides suitable for antagonism of PPIs. The CIS→PCA system described in Chapter 4 represents an attractive selection system, as it is able to sample an appreciable proportion of available peptide chemical space and restrict isolated hits to those stable in the bacterial cytosol. ***It was thus hypothesized that shorter peptides than those previously identified (32 – 37 amino acids) that nevertheless display attractive affinity for cJun could be selected by CIS→PCA.*** This approach intended to explore the minimum length for cJun interaction and identify shorter peptides with sufficient ability to inhibit AP-1 formation. Such peptides could then in the future provide novel alternative templates for helix constraint and truncation and other development strategies towards a cancer therapy.

This Chapter describes the use of CIS→PCA to select cJun binders from five peptide libraries. The first two libraries ("12mer" and "16mer" libraries) were

composed of 12 or 16 amino acids (three or four helical turns), with the aim of identifying helical or alternatively structured binders of cJun. The third library consisted of a two-and-a-half heptad library ("2.5h") based on the FosW LZ CC domain and was analogous to the 3.5h library of Chapter 4, similarly built by the ProxiMAX technique (Ashraf *et al.*, 2013). The final two libraries were peptides of three-and-a-half heptads in length; the first was the 3.5h ProxiMAX library of Chapter 4, which in this Chapter was screened with an altered screening strategy to try to overcome some its length issues, whilst the second was a library with alternative residue randomisations to 3.5h, that was constructed using an oligonucleotide primer PCR approach ("3.4hFosW") and therefore avoided ProxiMAX length complications. CIS→PCA was performed as in Chapter 4, and peptides were selected whose ability to bind cJun was predicted *in silico* and characterised *in vitro*. The results presented here suggest that CIS→PCA selection at the selection stringencies described was unsuccessful in selecting peptides with desirable affinity for cJun whilst being of a shorter and more therapeutically attractive size than longer antagonists such as the CPW peptide of Chapter 4. Thus, further efforts will be necessary to identify short AP-1 antagonists with more optimal binding properties.

5.1.1. Experimental Approach

The objectives of this work were as follows. Firstly, 2.5h and 3.4hFosW libraries were designed (Figure 2.1). The 2.5h library was built by ProxiMAX and was analogous to the 3.5h library design (as described in Chapter 4) but one heptad shorter, again hypothesizing that a broad range of residues with diverse physicochemical properties would provide the greatest opportunity for selection of

those affording appreciable enthalpy and binding free energy in short peptides. The 3.4hFosW library was designed to more strictly probe three-and-a-half heptad length peptides, and to investigate an even broader range of residues at ***a***, ***e*** and ***g*** residues, whilst fixing ***d*** positions as Leucines and with lesser diversity at ***b***, ***c*** and ***f*** positions. This design was based on the former four positions contributing strongly to CC formation preference and stability (Mason and Arndt, 2004), and fixing Leu residues to promote dimerisation (Harbury *et al.*, 1993). To construct the 12mer, 16mer and 3.4hFosW libraries, a library oligonucleotide was PCR amplified onto CIS display DNA constructs including the RepA-encoding gene to create a promoter-library-RepA construct ready for CIS display. CIS display screening of all libraries were performed as described in Chapter 4, including use of the same quantities of cJun in the selection for "low" and "high" selection stringencies. PCA screening was also performed as in Chapter 4, with the exception of the rescreening of the 3.5h library, which only screened peptides of the correct length from the CIS display enriched peptide subset (see 5.2.3.2). Selected peptides were analysed *in silico* for prediction of helicity and CC interaction stability, and *in vitro* using CD spectroscopy and ITC as described in 3.2.2 and 4.1.1 with peptides dissolved in KPP buffer unless stated otherwise.

5.2 Results

5.2.1 Library designs

5.2.1.1 Design of the 12mer and 16mer libraries

12mer and 16mer libraries (Figure 2.1) were designed to be fully randomised with all 20 naturally occurring amino acids at each position, to result in an even greater diversity of residues than the 2.5h library (see 5.2.1.2). Peptides from these

libraries may or may not form stable helices: as such, residue options were not restricted to only those found at high frequencies in CC domains to allow the opportunity for peptides with non-CC binding modes but appreciable affinity to also be selected. Further, this results in the broadest possible range of potential enthalpic contacts with cJun residues, such that combinations which provide strong binding are more likely to be identified. This randomisation also covers the entire range of possible peptide helicities. Helicity is aided slightly by an additional highly helical Ala residue following the translation initiator Methionine N-terminal of all library peptides in the CIS display construct, itself acting as a slightly stabilising N-terminal cap residue (Doig and Baldwin, 1995). Libraries were encoded by NNN codons, where N represents A/C/G/T nucleotides, thus encoding for all amino acids including both Cys codons and all three stop codons. This had the disadvantage of undesirable possibilities for library peptide disulfide crosslinking which may preclude binding to cJun and result in peptide removal by washing steps in CIS display, and undesirable premature truncation which would abolish RepA expression, both possibilities which decrease peptide diversity actually screened. However, it is was anticipated that functional peptides could nevertheless be selected from the remaining diversity.

5.2.1.2 Design of the 2.5h library

The 2.5h library (Figure 2.1) was designed to be a two-and-a-half heptad re-randomised version of FosW based on similar principles as the 3.5h library of Chapter 4 (see 4.2.2). In brief, this featured canonical core hydrophobic residue options (Mason and Arndt, 2004), with the addition of aromatic residues Phe, Tyr and Trp at 5 % total frequency to probe their effects/provide opportunity for higher enthalpy contacts perhaps necessary in short peptides. Prototypical charged and polar

residues were included at ***e*** and ***g*** positions, and charged, polar and hydrophobic residues were included at ***b***, ***c*** and ***f*** positions as observed in nature (Mason and Arndt, 2004, Crooks, 2013) and to probe the function of hydrophobic residues at these positions. It was expected that selected peptides may bind with their C-termini aligned with that of cJun. This would represent truncation of FosW from the N-terminus by two heptads, based on the rationale that the helicity of FosW is mainly concentrated at its C-terminus (see Figure 4.1 and 4.2.2.1). This was expected to be the preferred binding site based on considerations of the length of the linker between peptide and mDHFR in PCA screening. Alternatively, the 2.5h library options included an Asn at ***a***2 which could form Asn–Asn contacts with the cJun ***a***3 Asn, such that peptides spanned the centre heptads of cJun when bound. This was intended to probe truncation of FosW from fraying N- and C-termini (Rohl *et al.*, 1992), which may therefore feature the weakest enthalpic contacts: 2.5h peptides binding instead to the centre heptads where contact distances are more optimal may afford peptides the best binding free energy. This also allowed use of the same heptads as for 3.5h construction. Finally, as for 3.5h, AS and GAP helix-capping motifs were used to further stabilise peptide helices (Doig and Baldwin, 1995).

5.2.1.3 Design of the 3.4hFosW library

The 3.4hFosW library was designed as an alternative to 3.5h and probed different residue randomisations (Figure 2.1), both through positive design choices and less desirable but unavoidable coding of certain amino acids due to the use of degenerate DNA codons, a disadvantage that prompted the use of ProxiMAX (which avoids this issue) for 2.5h and 3.5h library construction (Ashraf *et al.*, 2013).

At core ***a*** positions, Ile, Leu and Val were included as typical choices for van der Waals' contacts, as was hydrophobic Met. Polar Asn and Gln residues provided opportunities for higher specificity interactions than hydrophobe–hydrophobe contacts, in particular with the ***a***3 Asn of cJun (O'Shea *et al.*, 1991, Gonzalez *et al.*, 1996). Positively charged Lys and His residues provided opportunity for specification of a dimeric CC interaction, either alone or potentially via interaction with ***g*** position residues (McClain *et al.*, 2002, Campbell and Lumb, 2002). His also probed aromatic selection at ***a*** positions as in the 3.5h library. Asp and Glu were included as negatively charged alternatives to Lys and His for potential ***a***–***g*** interactions. Core ***d*** positions were fixed to Leu to promote parallel dimerisation with cJun (Harbury *et al.*, 1993).

Positions ***e*** and ***g*** (except ***e***3 and ***g***4) featured a broad range of residue types. Choices capable of interhelical interactions included charged Asp, Glu and Lys, and polar Asn and Gln, whilst other choices were provided where interhelical interactions were not possible due to lack of electrostatic partner. These included Alanine for helicity induction, polar options Ser and Thr for solubility, hydrophobic options Ile, Leu, Val and Met to pack against the core and shield it from solvent, and aromatics Phe, His and Tyr as residues rarely screened at ***e*** and ***g*** positions to see if they would be selected for core shielding. Finally, Pro and stop codon options were undesirable but unavoidable due to degenerate codon use: however, Pro is unlikely to be selected in helical peptides due to helix-breaking tendency (O'Neil and Degrado, 1990), and severely truncated peptides resulting from stop codons should not be selected as this prevents RepA protein expression required for library peptide display to the target in CIS display selections. At ***e***3 an even broader range of residues including Arg and Trp explored whether an ***e***3–***g'***2 interhelical interaction – not present in the FosW–cJun

crystal structure (Chapter 3) – could be achieved with the ***g'***2 Gln of cJun, or offer a variety of amino acids for packing against the core. Undesirably, this also added Cys and Gly as options, though Gly is likely to be disfavoured as it is too conformationally flexible for inclusion in α-helices (O'Neil and Degrado, 1990), and peptides containing Cys as mentioned above are likely to be removed during selection under non-reducing conditions. Position ***g***4, on the other hand, featured a simpler set of options as there was no partner cJun residue for ***g***4–***e'***4.5 interaction, and included Ala, Glu, Lys and Thr to cover helical, charged and polar residues. Finally, at outerface ***b***, ***c*** and ***f*** positions, Ala, Glu, Lys and Thr were probed for the same reasons. As for the 3.5h library, the 3.4hFosW library was designed with the expectation for binding to cJun such that library peptide and cJun C-termini aligned.

5.2.2 Library construction, QC and diversities

12mer, 16mer and 3.4hFosW libraries were created by stepwise PCR amplifications, involving library degenerate oligonucleotide addition to the RepA CIS display construct and then addition of the *tac* promoter to the library 5' end. Generation of the desired amplification products was evidenced by observation of bands of the expected molecular weights on agarose gels: example amplifications for 12mer and 16mer libraries are shown in Figure 5.1. For the 2.5h library, construction involved ligation of Heptad cassettes 1, 2 and 4.5 (see Chapter 4 Appendix Figure 4.1, 4.2 and 4.3) to create a Heptad 1-2-4.5 construct, which was ligated to the RepA CIS display construct and amplified before final addition of the *tac* promoter via PCR. Figure 5.2 displays DNA constructs generated during each stage of the 2.5h build run on an agarose gel, which again confirmed generation of the desired products.

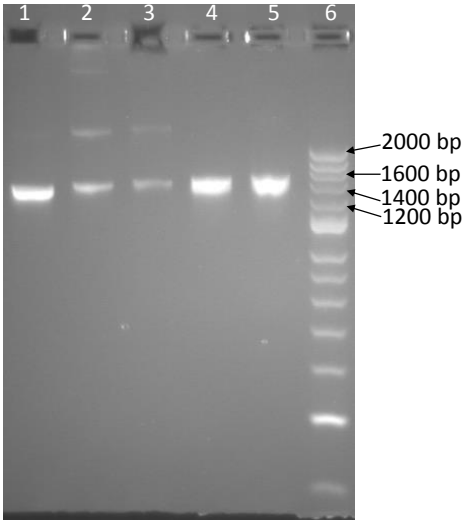


Figure 5.1: 12mer and 16mer library builds. Samples and ladders run on 2% agarose/Tris-Borate EDTA gel stained with SybrGold. Expected sizes of desired species are given in parentheses. Lane 1: link-RepA-CIS-Ori (1339 bp), Lane 2: 12mer-link-RepA-CIS-Ori (1380 bp), Lane 3: 16mer-link-RepA-CIS-Ori (1392 bp), Lane 4: *tac*-12mer-link-RepA-CIS-Ori (1460 bp), Lane 5: *tac*-16mer-link-RepA-CIS-Ori (1472 bp), Lane 6: Hyperladder II (Bioline).

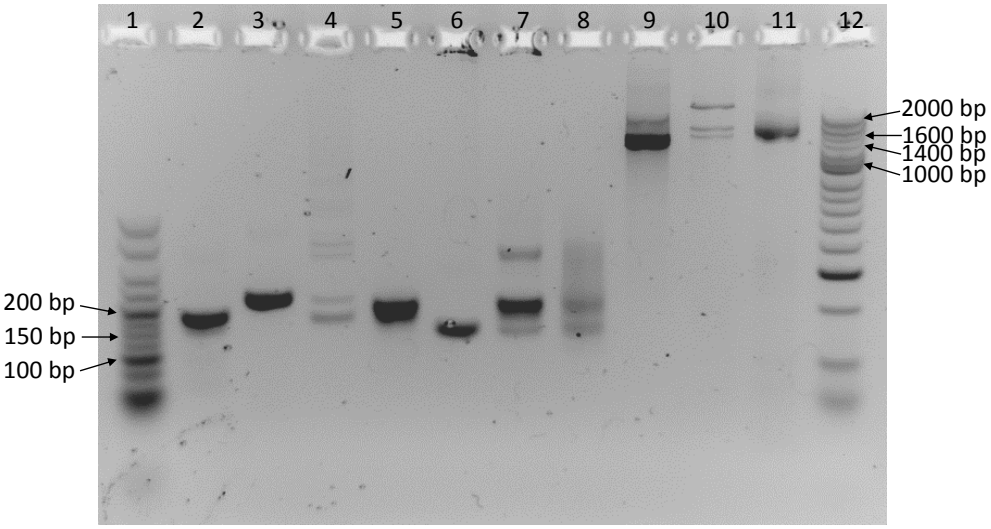


Figure 5.2: 2.5h library build. Agarose gel of stages of the 2.5h build, including Heptad ligations, digestions, and the final PCR of CIS-display ready construct, *tac*-2.5h-link-RepA-CIS-Ori. Samples and ladders run on 2% agarose/Tris-Borate EDTA gel stained with SybrGold. Expected sizes of desired species are given in parentheses. Lane 1: Hyperladder V (Bioline), Lane 2: Heptad 1 digest (173 bp), Lane 3: Heptad 2 digest (212 bp), Lane 4: Heptad 1-2 ligation, Lane 5: Heptad 1-2 digest (194 bp), Lane 6: Heptad 4.5 digest (143 bp), Lane 7: Heptad 1-2-4.5 ligation, Lane 8: Heptad 1-2-4.5 digest (206 bp), Lane 9: (MluI)link-RepA-CIS-Ori digest (1344 bp), Lane 10: Heptad 1-2-4.5-link-RepA-CIS-Ori ligation, Lane 11: tac-1-2-4.5-link-RepA-CIS-Ori PCR (1478 bp), Lane 12: Hyperladder II (Bioline).

Library sequencing was performed to assess the quality of constructed libraries and consequent functional peptide diversity. For 12 and 16mer libraries, considering the well characterised library qualities produced using the Isogenica Ltd. (Little Chesterford, UK) Standard Operating Procedures followed during library construction, the simple two-step PCR construction, and the fully randomised nature of these libraries, library sequencing was not repeated prior to selections. Instead, analysis of post-selection sequencing indicated that 89 – 96 % of sequences selected were the expected length, as would be expected for these libraries. For the 3.4hFosW library, the similar PCR construction technique additionally made use of high-fidelity KOD polymerase to limit PCR errors, and post-selection sequencing similarly identified 88 – 93 % of sequences to be the expected length. The completed 2.5h variable region (Heptad 1-2-4.5) was deep sequenced to assess library quality. From a random sample of 91,961 sequences, observed codon incorporation frequencies of peptides of the expected length (Figure 5.3) were very close to those expected from the library design (Figure 2.1), and the orientation of Heptad cassettes was also as expected, suggesting that peptides of the correct randomisations had been constructed using ProxiMAX. Following appendage of the RepA CIS display construct to the 2.5h library variable region, Sanger sequencing confirmed the expected RepA sequence (proprietary, data not shown).

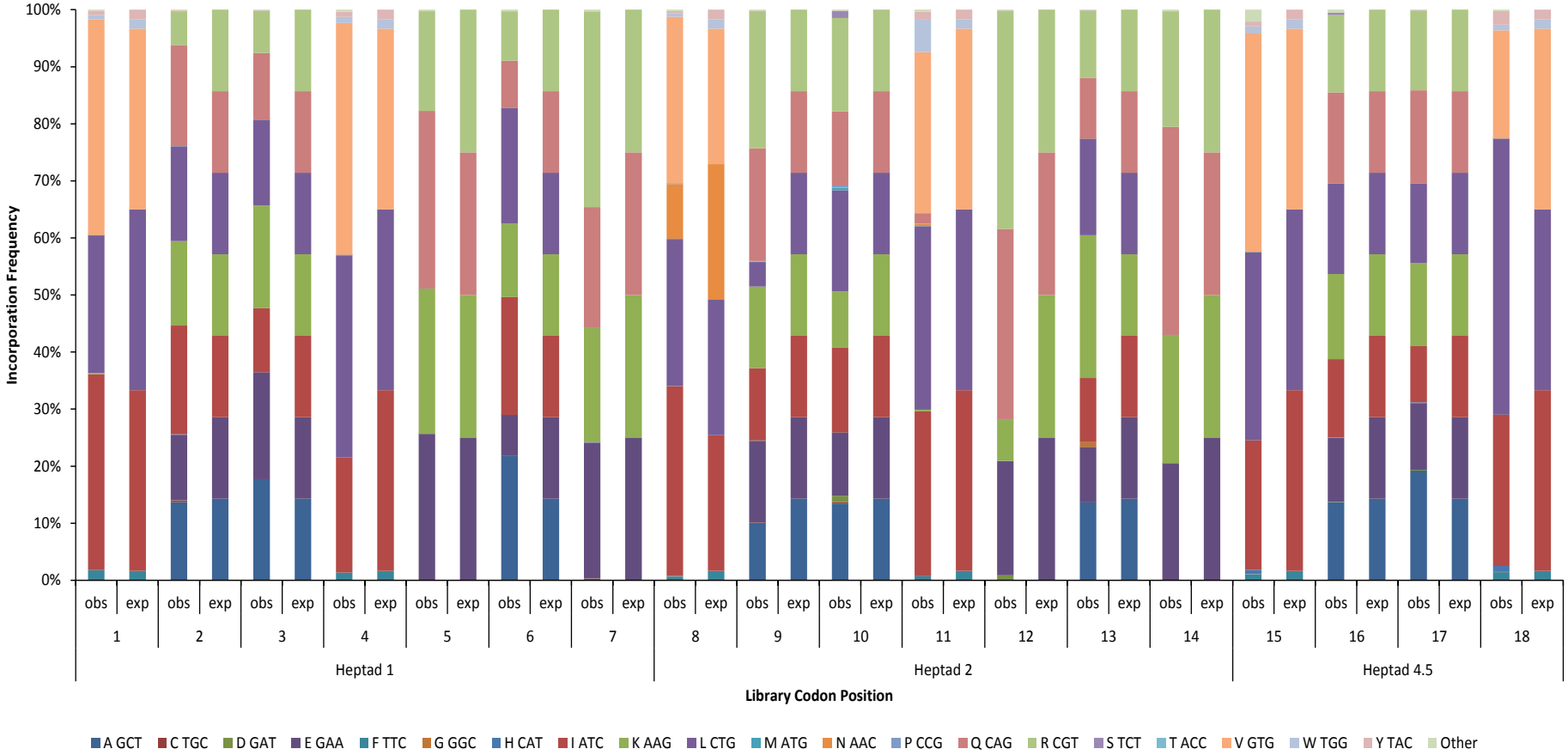


Figure 5.3: Deep sequencing QC of codon incorporation frequencies observed vs. expected for the completed 2.5h library variable region. Observed (obs) and expected (exp) incorporation frequencies are shown for library amino acids/codons (coloured) at positions 1 – 18 of correct length peptides of the completed 2.5h variable region (Heptad 1-2-4.5).

However, as for the 3.5h library, deep sequencing identified peptides with lengths other than the 18 residues desired (see also 4.2.3). Correct length peptides made up 70 % of the library, in closer agreement than the 3.5h library with the percentage of correct length peptides reported previously for a 21mer ProxiMAX-built library (75 %) (Frigotto *et al.*, 2015). Of the remaining constructs, 6 % were longer than desired, and 24 % were shorter. Of the shorter sequences, 6% featured single nucleotide deletions, 2 % dinucleotide deletions, 4 % codon deletions, and 4 % two-codon deletions. As described in Chapter 4, longer peptides in the library may be preferentially selected over correct length peptides as longer peptides are more likely to have higher affinity for the target through a greater net interaction enthalpy (in addition to entropy effects). The percentages of shorter and longer sequences in the 2.5h library were approximately one-third and one half of those for the 3.5h library. This could mean that issues of peptide length were less significant for the 2.5h library; however, as peptide length decreases from 3.5h to 2.5h and affinity is predicted to decrease, longer peptides may present greater competition to 2.5h peptides than to 3.5h peptides. Nevertheless, as argued for the 3.5h library, such length issues did not necessarily preclude selection of correct length peptides, and additionally considering that 2.5h library peptides (18mers) are significantly shorter than previously derived cJun binders like 37mer FosW, selection of peptides between 18 and 37 residues would still achieve the aim of selection of a shorter cJun binder than FosW. In order to encourage selection specifically of correct length peptides, steps were taken during PCA selection of this library (as well as for the 3.5h library) and are described in 5.2.3.2.

Theoretical library diversities are presented in Figure 2.1, excluding stop codons to represent only the number of peptides that could be presented to cJun in CIS display. The diversity of all four libraries actually screened in CIS display was calculated to be ≈6.1 x 10^{10} peptides (see section 2.5 for calculation). Of this diversity, the issues with peptide length for 2.5h meant that 70 % of peptides screened (i.e. 4.3 x 10^{10}) were of the correct length. For all libraries, the diversity screened was an appreciable proportion of the theoretical diversities, and was much higher than in previous selections against cJun where ≈1.2 million-fold lesser diversity was screened (Mason *et al.*, 2006, Crooks *et al.*, 2011).

5.2.3 CIS→PCA selections

CIS display was used to screen for cJun binding peptides in 12mer, 16mer, 2.5h, 3.5h (described in 4.2.4) and 3.4hFosW libraries using the same "low" and "high" stringency conditions as in Chapter 4 (see 4.3.1). CIS display was followed by PCA screening of peptides selected by both CIS display stringency selections except for 12mer and 16mer libraries, and the 2.5h high stringency condition. In addition, a modified PCA protocol was applied to 2.5h and 3.5h CIS output peptides to encourage correct length peptide selection.

5.2.3.1 CIS display of 12mer and 16mer libraries

Deep sequencing identified ≈200,000 – 325,000 unique sequence reads (depending on sample) after CIS display selection of the 12mer and 16mer libraries, of which 10 – 411 peptides were considered to be 'enriched' (≥50 observations) and were analysed further. The top 20 peptides identified by sequencing and ranked by observation frequency are given in Appendix Table 5.1 and 5.2, and the overall

composition of lower ranked sequences are given in Appendix Table 5.5. Observation frequencies of low and high stringency 12mer and 16mer peptides (generally ≤700 observations) (Figure 5.4) were much lower than expected frequencies representative of enrichment, such as those of 3.5h peptides for example (≈20,000 observations by R4), though it is noted that observation frequencies are dependent on many factors and so any threshold frequency chosen to indicate enrichment is somewhat arbitrary. Further, whilst observation frequencies of the few most frequently observed peptides do increase between R3 and R4, which could indicate specific enrichment of these sequences, so too does the number of peptides observed ≥50 times, resulting in similar or decreased relative frequencies of peptides from R3 to R4. This suggests that enrichment of specific peptides was very inefficient, allowing PCR errors to create more 'unique' sequences, increasing diversity. Of the two selection stringencies, the low stringency conditions appeared to have enriched peptides slightly better than high stringency conditions, with some separation between the most frequently observed peptides and peptides with relative frequencies ≤5 %, but generally observation frequencies were too low to suggest efficient enrichment. The 16mer selections appear to have been slightly more efficient than the 12mer selections, but still were not as efficient as expected. In light of this, and *in silico* analysis of the most frequently observed peptides suggesting poor biophysical properties (see 5.2.4), 12mer and 16mer library output peptides were not cloned into PCA for continued screening.

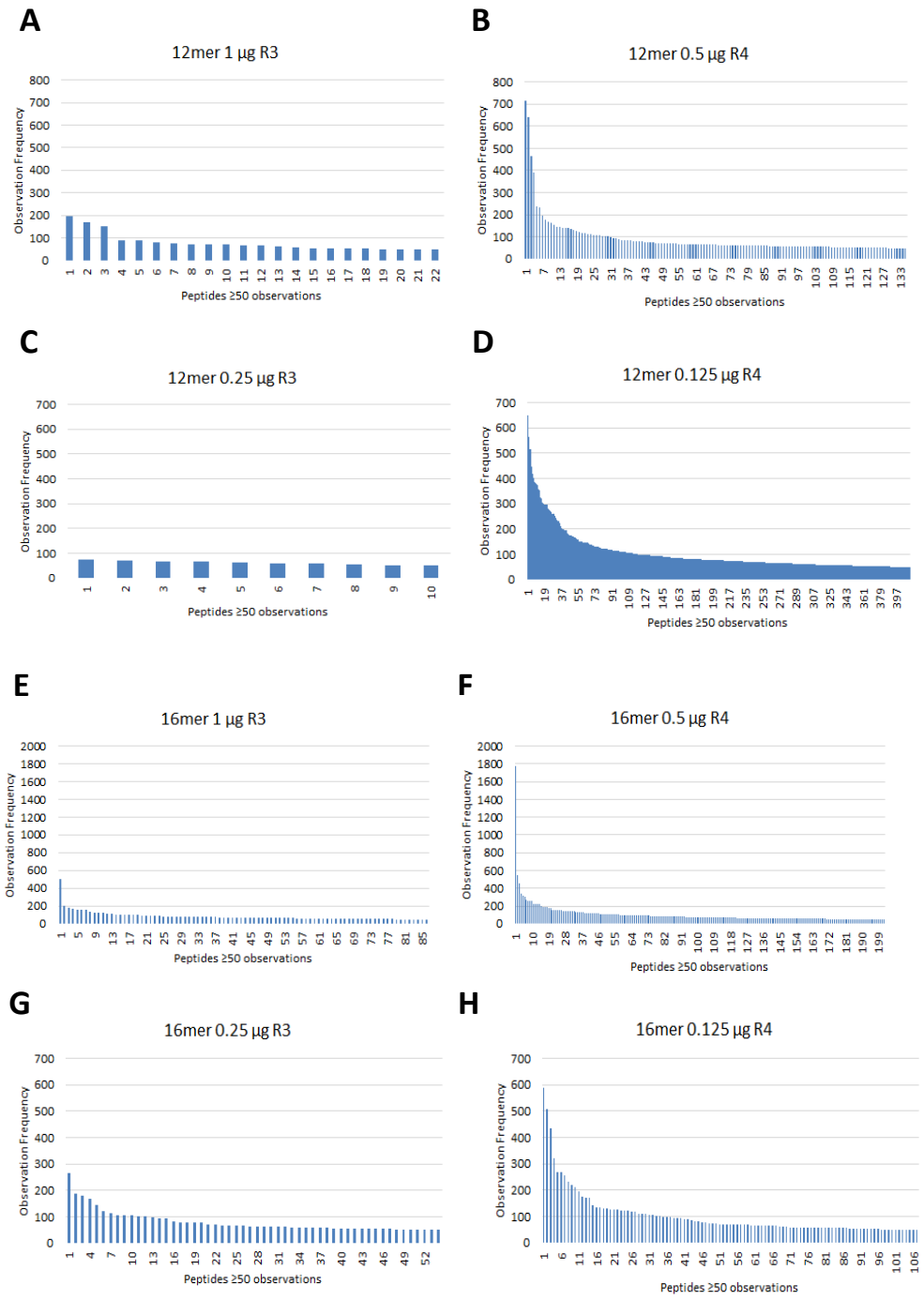


Figure 5.4: Poor enrichment of peptides from the 12mer and 16mer libraries by CIS display. Observation frequencies of peptides observed ≥50 times in the peptide subsets selected by CIS display, as identified by deep sequencing. Peptides ranked in order of observation frequency. **A** and **B**; peptides from the 12mer low stringency selection round R3 and R4. **C** and **D**; peptides from the 12mer high stringency R3 and R4. **E** and **F**; peptides from the 16mer low stringency selection R3 and R4. **G** and **H**; peptides from the 16mer high stringency selection R3 and R4. cJun quantities determining selection stringency are given in µg (see Table 2.11 for further explanation).

5.2.3.2 CIS→PCA selection of the 2.5h library

Observation frequencies for 2.5h library peptides selected by CIS display are shown in Figure 5.5A-D (see Appendix Table 5.3 and 5.5 for the top 20 peptides and the overall composition of remaining lower ranked sequences). Deep sequencing identified a total of 51,981 – 69,884 unique sequence reads (depending on sample) after CIS display, of which 223 – 630 peptides were considered enriched. In the low stringency selection, at R3 57 % of enriched peptides were the correct length and at R4 this increased to 65 %, whilst in the high stringency selection correct length peptides made up 65 % at R3 and R4, such that the majority of selected peptides were the expected length in both selections. The observation frequencies of the most frequently observed peptides (≈1600 – 2500 observations) are slightly higher than those of 12mer and 16mer selections (generally ≈700 observations) which could indicate more efficient selection of 2.5h peptides, but are still below those expected for an efficient selection. Further, as for 12mer and 16mer libraries, there was an increase in the number of enriched peptides from R3 to R4, and a consequent decrease in peptide relative frequencies, to relative frequencies lower than for 12/16mer peptides. Based on these analyses, the 2.5h library selection had not efficiently selected cJun-binding peptides.

To investigate whether valuable cJun binders had nevertheless been enriched by CIS display, the low stringency enriched peptide subset (as the most likely selection to have been successful based on arguments made in Chapter 4) was cloned into the pET28a+ PCA vector for PCA screening. CIS display R4 deep sequencing identified 51,981 unique sequence reads for this stringency, the diversity of which was covered to 94.5 % in the subsequent PCA according to Equation 1 (see 2.4.3).

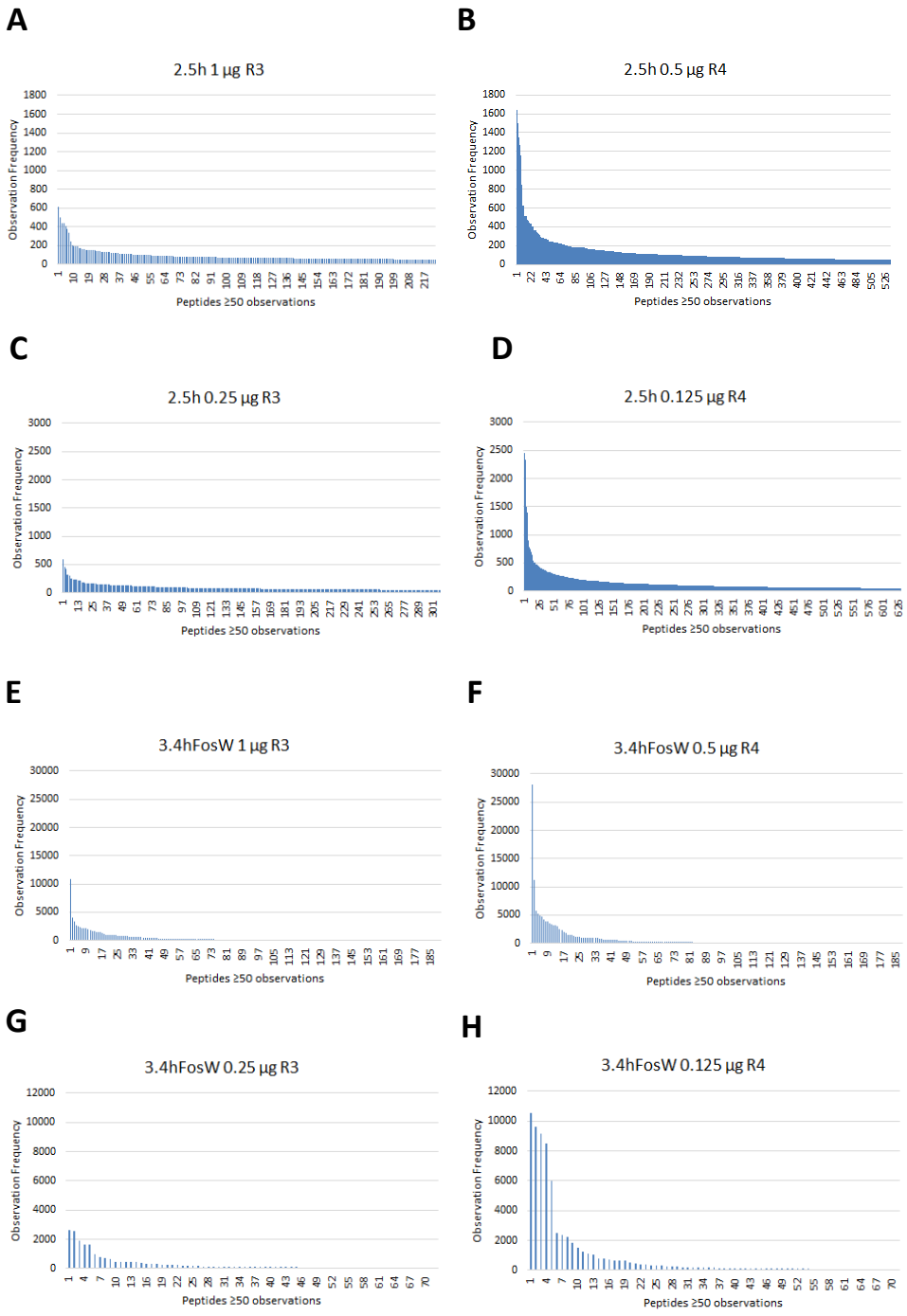


Figure 5.5: Poor enrichment of peptides from the 2.5h library and some enrichment of peptides from the 3.4hFosW library by CIS display. Observation frequencies of peptides observed ≥50 times in the peptide subsets selected by CIS display, as identified by deep sequencing. Peptides ranked in order of observation frequency. **A** and **B**; peptides from the 2.5h low stringency selection round R3 and R4. **C** and **D**; peptides from the 2.5h high stringency R3 and R4. **E** and **F**; peptides from the 3.4hFosW low stringency selection R3 and R4. **G** and **H**; peptides from the 3.4hFosW high stringency selection R3 and R4. cJun quantities determining selection stringency are given in µg (see Table 2.11 for further explanation).

PCA was performed as described in 2.7. However, no colonies were observed on M9 agar using the normal concentration of trimethoprim antibiotic. This suggested that peptides could not bind cJun with high enough affinity for sufficient mDHFR reconstitution and activity to allow colony growth at this trimethoprim stringency (Remy and Michnick, 1999). At this stage, an altered PCA was performed to investigate whether any 2.5h peptides had affinity close to (i.e. just below) the threshold applied by the trimethoprim stringency, and that could represent attractive peptides given their much shorter length to previous antagonists. This PCA (named "correct-length PCA") allowed Competition Selection only of peptides of exactly the correct length, to counter length issues described in 5.2.2, and was further used to rescreen the 3.5h library of Chapter 4 (see 5.2.3.3).

Correct-length PCA involved plating of PCA cells on M9 agar (with antibiotics and IPTG) that contained only half the normal concentration of trimethoprim. This concentration was chosen based on the relative lengths and expected affinities of 4.5-heptad peptides (e.g. FosW, screened with 100 % trimethoprim), 3.5-heptad peptides (i.e. 3.5h library peptides) 78 % of FosW's length and so screened on 75 % trimethoprim in correct-length PCA (see 5.2.3.3), and 2.5-heptad peptides (i.e. 2.5h library peptides) which were 56 % of FosW's length and so were screened on 50 % trimethoprim. Sanger sequencing of extracted library plasmids was used to identify which colonies contained correct length peptides. Only these colonies were picked into liquid LB cultures with plasmid-selecting antibiotics, before an equal volume of optical density-normalised LB culture (OD_{600} = 0.2 AU) was inoculated into liquid M9 for Competition Selection, again using 50 % trimethoprim.

For correct-length PCA of the low stringency 2.5h CIS selection output peptides, 28 colonies were observed on the Single Step selection plate. Sequencing of these identified a majority of 18 peptides of the expected 18 residue length (sequences in Appendix Table 5.6), two shorter peptides, five peptides 20 – 23 residues long and three poor sequencing reads/mixed colonies. Interestingly, the 18 correct-length peptides included peptides at positions 1 and 4 in the CIS display output deep sequencing, as well as peptides ranging down to rank position 465 after CIS display. These 18 peptides were taken forward into Competition Selection, where sequencing of the P3 passage pool (Appendix Figure 5.1) identified a clean sequence corresponding to the most frequently observed peptide after CIS display, hereafter referred to as “2.5pep1” (Table 5.1).

Table 5.1: CIS display and CIS→PCA selected peptides from the 12mer, 16mer, 2.5h, 3.5h and 3.4hFosW libraries are not predicted to interact with cJun with high affinity. Displayed are peptide observation frequencies, and relative observation frequencies as a percentage of all observed sequences, from sequencing of CIS display outputs. Values for peptides CPW from Chapter 4, and previously derived FosW, 4hFosW, and $FosW_{CANDI,}$ and native Fos and Jun homologues are provided for comparison. Peptide helicity was calculated by Agadir (Lacroix *et al.*, 1998), predicted interaction T_m values were calculated by bCIPA (Mason *et al.*, 2006), and predicted aggregation propensities were calculated using PASTA2.0 (Trovato *et al.*, 2006) and Zyggregator (Pawar *et al.*, 2005). *AS and GAP helix capping motifs added for Agadir, PASTA 2.0 and Zyggregator analyses. †Where two bCIPA T_m values are given, this relates respectively to N- and C-terminal alignment of the peptide to its longer cJun partner; the more positive T_m indicates the preferred binding site. ‡Sequences of native cJun binders from Angel and Karin (1991). Values are given to 2 s.f. except observation frequency (raw number). Low predicted aggregation propensity is indicated by a more positive PASTA 2.0 score, or a more negative Zyggregator score, and via comparison with scores of known soluble peptides cFos (-3.0/-6.8), FosW (-1.5/-7.3) and cJun (-1.7/-5.5) (PASTA 2.0 score/Zyggregator score respectively).

Library and selection stringency	Position in CIS selection	Peptide name	Peptide sequence	Observation frequency	Relative observation frequency (%)	Helicity (%) (Agadir) *	Predicted interaction T_m (°C, bCIPA)†	Predicted aggregation propensity score (PASTA 2.0/Zyggregator algorithms)*
12mer CIS low stringency	1st		SYDVTITITLKM	715	5.7			-7.0/-0.74
12mer CIS high stringency	1st		YHIKIEVNFHFP	648	1.3			-4.5/-1.8
16mer CIS low stringency	1st		KKINMFYTEVYIVIIE	1772	8.3			-11/-4.0
16mer CIS high stringency	1st		KFTFTIIMQFNNHYFL	589	5.3			-7.5/-0.9
2.5h correct-length CIS→PCA low stringency	1st	2.5pep1	VKAVREQLIIIQIKLLEL	1645	2.3	17	-23/-46	-7.4/-2.6
3.5h correct-length CIS→PCA low stringency	Not observed	3.5pep22	IKKIELQLLEKRIQIVRLKIQI	Not observed	Not observed	10	-17/-55	-5.2/-3.7
	Not observed	3.5pep23	IKRLEVRIIEKRIQIVRIQVRIV	Not observed	Not observed	2.8	-23/-61	-9.0/-3.9
3.4hFosW CIS low stringency	1st	3.4LCIS	LTALIEQNAALLETLKTLSTEIEKL	28198	20	4.0	-9/-11	-3.1/-5.6
3.4hFosW CIS high stringency	1st	3.4HCIS	LKKLLEYLETLSAHVTTLHATVEKL	10552	15	10	-26/-23	-5.7/-4.8
3.4hFosW CIS→PCA low stringency	9th	3.4W	LTALIEQNTALLELLKALSTELEKL	3896	2.7	16	-16/-8	-3.1/-5.7
3.5h CIS→PCA low stringency selection	71st	CPW	VQEIEQEIQELEKRIKQIQQEFQEIEQQIALL			15	53	-3.6/-7.0
PCA		FosW	LDELQAEIEQLEERNYALRKEIEDLQKQLEKL			26	63	-1.5/-7.3
PCA		4hFosW	LDELQREIEQLEELNYALQKEIEDLQKQ			22	32/-45	
PCA		$FosW_{CANDI}$	LDELQAEIEQLEDQNYALQKEVEDLRKELEKL			22	57	
Native		cFos‡	TDTLQAETDQLEDEKYALQTEIANLLKEKEKL			3.5	13	-3.0/-6.8
Native		FosB‡	TDRLQAETDQLEEEKYELESEIAELQKEKERL			5.0	14	
Native		Fra1‡	TDFLQAETDKLEDEKYGLQREIEELQKQKERL			14	12	
Native		Fra2‡	TEKLQAETEELEEEKYGLQKEIAELQKEKEKL			4.1	18	
Native		cJun‡	IARLEEKVKTLKAQNYELASTANMLREQVAQL			3.7	7.0	-1.67/-5.49
Native		JunB‡	IARLEDKVKTLKAENYGLSSTAGLLREQVAQL			2.4	-4.0	
Native		JunD‡	ISRLEEKVKTLKSQNYELASTASLLREQVAQL			2.6	2.0	

5.2.3.3 CIS→PCA selection of the 3.5h library

Following CIS display selection of the 3.5h library (see 4.2.4.1), correct-length PCA was performed on CIS output peptides using trimethoprim at 75 % normal concentration. For the low stringency CIS selection carried forward into correct-length PCA, again covering >99 % of unique sequence reads as for the original PCA (see 4.2.4.3), 17 colonies were observed on the Single Step selection plate. Sequencing of these identified two peptides shorter than the expected 25 amino acids, 14 longer peptides (27 – 38 amino acids), and one mixed read (sequences in Appendix Table 5.7), and so was not as successful as the 2.5h correct-length PCA Single Step selection in identifying correct-length peptides to take forward into Competition Selection. The longer peptides included four copies of the most enriched peptide after CIS display in the low stringency 3.5h selection ("CIS1" in Chapter 4). Neither of the two shorter peptides, a 22mer and a 23mer, were observed in the 25,884 peptides identified by low stringency R4 deep sequencing. These peptides, hereafter referred to as "3.5pep22" and "3.5pep23" (Table 5.1), were not taken forward into Competition Selection, and instead were characterised as desirably shorter peptides to determine whether they had appreciable affinity for cJun.

For the high stringency selection, again covering 95 % of sequenced peptides as in the normal PCA (see 4.2.4.3), two Single Step selection colonies grew very quickly and were not followed by others in the timeframe the plates were incubated before picking and sequencing. These colonies contained a 32mer and 28mer not previously observed in deep sequencing of the high stringency CIS output. However, as no correct-length or near-correct length peptides were identified, no addition peptides were taken forward for characterisation.

5.2.3.4 CIS→PCA selection of the 3.4hFosW library

The 3.4hFosW library was also subjected to CIS→PCA. After CIS display, 44,392 – 119,434 unique sequence reads were recorded, of which observation frequencies of the 71 – 190 enriched peptides are displayed in Figure 5.5E-H (see Appendix Table 5.4 and 5.5 for summary of CIS output sequences). In contrast to 12mer/16mer and 2.5h selections, observation frequencies of the most enriched peptides for both stringencies were high (10,000 – 25,000), comparable to the 3.5h selections, suggesting successful enrichment of specific binders (though the number of enriched sequences did not decrease from R3 to R4, contrary to expectations).

For the low stringency selection, two sequences were enriched above the remaining 99 % of peptides of relative frequencies ≤5 %, to comprise 20 % and 8 % of the pool. For the high stringency selection, five sequences were separated from 93 % of the pool with relative frequencies between 15 % and 8 %. Thus, 3.4hFosW peptides appeared to have been selected for cJun binding capability. The top most enriched peptide from both the low and high stringency 3.4hFosW selections was synthesized for comparison with peptides selected by subsequent PCA, and are hereafter referred to as “3.4LCIS” and “3.4HCIS” (Table 5.1).

The suggested efficiency of CIS display enrichment prompted transfer of peptides to PCA. PCA colony numbers amassed (31,000 – 46,000) were lower than achieved for cloning of previous libraries, but were still sufficient to transfer the top 5,000 most abundant peptides that were sequenced as calculated according to Equation 1 (see 2.4.3), which was considered sufficient for further experimentation. Two colonies for the low stringency selection grew very quickly in the Single Step

selection and the corresponding peptides were sequenced, whilst the high stringency selection did not produce colonies within the time course of plate incubation. These two peptides had occupied positions 20 and 9 of the low stringency CIS output. Competition Selection between these peptides was performed to identify the higher affinity binder, of which the peptide previously at position 9 of the CIS display output and hereafter named "3.4Winner" ("3.4W") (sequence in Table 5.1), dominated the pools from P1 until then end of passaging at P3 (Appendix Figure 5.2), and so was taken forward for further characterisation.

5.2.4 *In silico* analysis of selected peptides

In silico algorithms were used to predict the likelihood of CIS and CIS→PCA selected peptides interacting with cJun and whether via CC or other binding modes. Peptides are compared to native binders and previous cJun antagonists in Table 5.1.

Firstly, Agadir reported predicted monomeric peptide helicities (Lacroix *et al.*, 1998). Peptide helicities were generally higher than native cJun binders, but below PCA-derived FosW, 4hFosW and $FosW_{CANDI}$, with 2.5pep1, 3.5pep22, 3.4HCIS, and 3.4W closer to that of CIS→PCA selected CPW from Chapter 4. These helicities nevertheless suggest good potential for binding cJun based on ability of CPW and PCA-derived peptides to bind cJun with high affinity, and native binders to at least bind cJun with affinities of relevance *in cellulo*. 2.5pep1, 3.5pep22, 3.4HCIS, and 3.4W additionally were surprisingly helical for their short length compared to FosW, 4hFosW and $FosW_{CANDI}$, especially 2.5pep1, which despite being the shortest peptide was also the most helical peptide from those derived in this Chapter.

Secondly, bCIPA analysis reported propensity of peptides to form stable CC interactions with cJun (Mason *et al.*, 2006), based on binding with N- or C-termini aligned with that of cJun. Unlike predictions for native and PCA-derived cJun binders (excluding 4hFosW), bCIPA T_m values for short peptides were all negative. This suggested either that peptides were unlikely to form CC interactions with cJun, or that bCIPA was unable to accurately predict interactions for peptides of this length. Calculation of BOW scores to compare with bCIPA analysis was not possible due to peptide lengths being below the cut-off for predictive accuracy (Fong *et al.*, 2004).

Thirdly, aggregation propensity algorithms PASTA 2.0 and Zyggregator were used to analyse aggregation potential as a proxy for preference of peptides for non-specific interactions (see also 4.2.5.3) (Trovato *et al.*, 2006, Pawar *et al.*, 2005). PASTA 2.0 predicted all short peptides to aggregate more readily than cFos, cJun and FosW negative controls except 3.4LCIS and 3.4W, whose aggregation propensities were similar to that of cFos and thus these peptides may be more specific binders. Zyggregator scoring (more negative scores representing lesser aggregation propensity) reports a similar ranking, with 2.5pep1, 3.5pep22, 3.5pep23 and 3.4HCIS all more aggregation prone than cJun, whilst other peptides were of similar aggregation propensities to controls. This suggested that 3.4W was likely the most specific peptide, and potentially that the CIS→PCA low stringency selection it was selected in may have been the most successful in selecting valuable cJun binders.

Aggregation propensity prediction was also performed on the top most enriched peptides from the 12mer and 16mer CIS display selections (Table 5.1) to corroborate suggestions of poor selection from low enrichment rates (see 5.2.3.1).

PASTA 2.0 and Zyggregator both score 12mer and 16mer peptides as much more aggregation prone than control cFos, cJun and FosW peptides, and these peptides are among the most aggregation prone compared to peptides selected from 2.5h, 3.5h and 3.4hFosW libraries, indicating that neither low or high stringency 12mer and 16mer CIS selections have been successful in enriching specific cJun binders.

5.2.5 Sequence alignment evidence of CC formation

Sequence alignments were performed to assess the similarity of 2.5pep1, 3.5pep22, 3.5pep23, 3.4LCIS, 3.4HCIS, and 3.4W to CC domains of cFos and FosW, compared to Fos homologues that align well to both cFos and FosW, and previously aligned peptides CIS1, CIS21 and CPW (see 4.2.5.4). Manual alignments were performed assuming that all peptides bound with C-termini aligned with that of cJun (Figure 5.6A and B): the lack of ***a***2 N selection in 2.5pep1 suggested alignment to the C-terminal of cJun as the expected binding mode (see also 5.2.1.2). Peptides generally aligned with reasonable similarities to cFos (57 – 73 %) and comparably to FosW (55 – 80 %). Overall the similarities were lower than those of CPW, CIS1 and CIS21 (69 – 84 % vs. cFos and 81 – 91 % vs. FosW) and Fos homologues, but are of sufficient similarity to cFos and FosW to suggest CC formation by short peptides. Furthermore, consensus sequences generated from short peptides in alignment with cFos and FosW resemble the generic CC protein consensus sequence n-p-p-n-p-p-p (Mason and Arndt, 2004), with minor deviations only at ***b*** and ***c*** positions which likely have minimal influence on interhelical interactions in the CC, or at ***e*** positions where hydrophobic residues may shield the core from solvent (O'Shea *et al.* 1991).

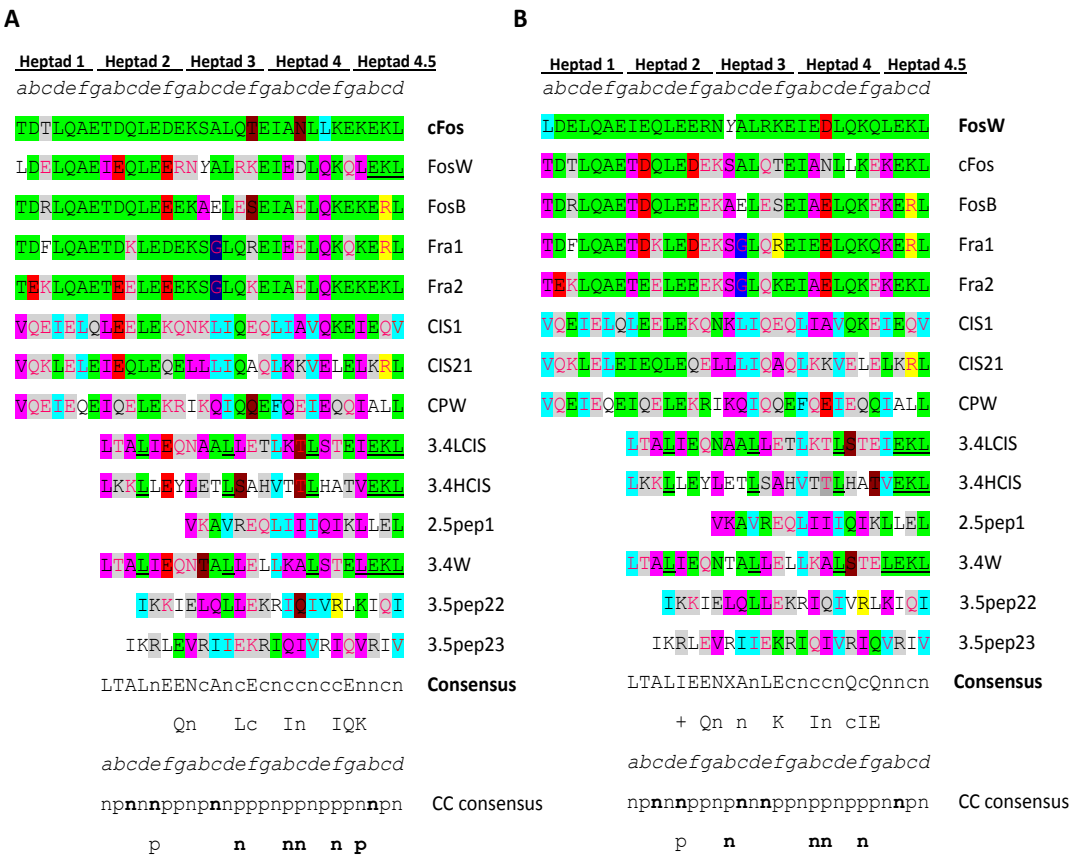


Figure 5.6: Similarity of CIS display selected 3.4LCIS and 3.4HCIS peptides, and of CIS→PCA selected 2.5pep1, 3.5pep22, 3.5pep23, and 3.4W peptides to known Jun/Fos CC proteins. Colours for residues in aligned sequences vs. the query sequence are as follows: green denotes identity, blue denotes hydrophobic similarity (Ala considered hydrophobic), yellow denotes positive charge similarity, red denotes negative charge similarity, brown denotes polar (uncharged) similarity, grey denotes polar/charged similarity, and pink (highlight or text colour) denotes residues different to that of the query sequence but identical across two or more peptides. Underlined residues were fixed in the library design. The Y in FosW is in italics to denote that it was not a residue selected from a library, and has not been aligned. Consensus sequences were derived from residue similarity colourings, except fixed residues which were not considered. For consensus sequences, n denotes non-polar residues, p denotes polar residues, + denotes positively charged residues, c denotes charged/polar residues, and X denotes non-polar/polar (any) residues. **A** and **B** Manual sequence alignments of peptides against query sequences cFos or FosW, with Fos homologues (and FosW for alignment to cFos) as reference alignments, based on C-terminal alignment of each peptide with cFos or FosW. Consensus sequences based solely on similarity between peptides 3.4LCIS, 3.4HCIS, 2.5pep1, 3.4W, 3.5pep22, 3.5pep23 and cFos or FosW have been constructed, and are compared to the coiled coil consensus n-p-p-n-p-p-p (Mason and Arndt, 2004), with positions not conforming to this pattern in bold text.

5.2.6 *In vitro* characterisation of cJun interaction

In vitro characterisation of short peptides by CD spectroscopy and ITC was performed as described in 3.2.2. CD and ITC data are compared against high affinity cJun antagonists FosW, $FosW_{CANDI}$, 4hFosW and native cFos–cJun AP-1 where CD and ITC data was available (Figure 5.7 and Table 5.2) (Mason *et al.*, 2006, Mason *et al.*, 2007b, Crooks *et al.*, 2011, Worrall and Mason, 2011). Peptides 2.5pep1, 3.5pep22 and 3.5pep23 were assayed by CD at 75 μM, 20 μM and 10 μM total protein concentration (including cJun for heteromeric mixtures) respectively due to solubility limits even in "low salt" buffer (10 mM KH_2PO_4/K_2HPO4), whilst 3.4LCIS, 3.4HCIS and 3.4W were assayed at 150 μM in KPP ("low salt" with 100 mM KF) buffer as normal. Raw CD ellipticity is reported as mean residue ellipticity (MRE) which is normalised for concentration and peptide length. Appendix Figure 5.3 demonstrates that cJun data recorded across the experimental range of concentrations and in both buffers are comparable by reporting MRE, such that heteromeric traces can be compared across all peptides.

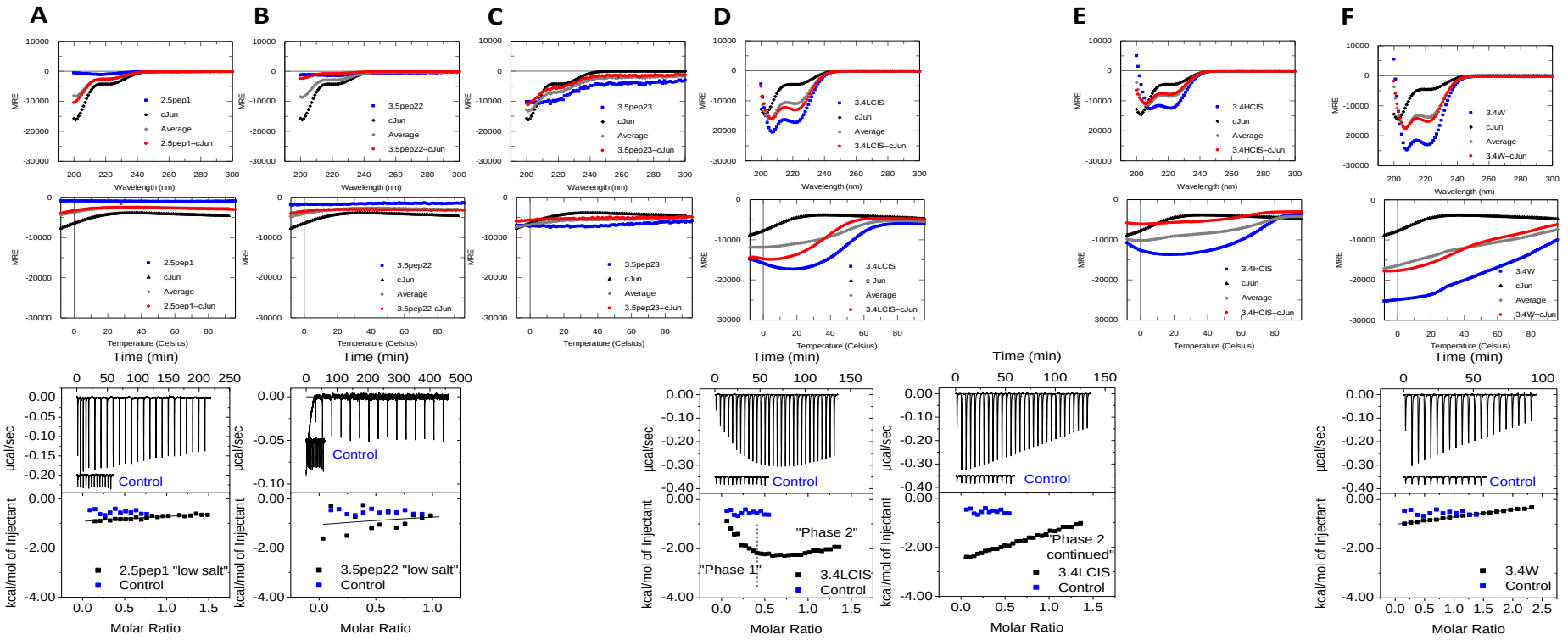


Figure 5.7: CIS display and CIS→PCA selected peptides bind cJun with low affinity (legend continues on next page). Biophysical characterisation of cJun interaction by circular dichroism spectroscopy (CD) scans (top panels) and thermal denaturation (middle panels), and by isothermal titration calorimetry (ITC) (bottom panels, not all peptides analysed by ITC). Peptides 2.5pep, 3.5pep22 and 3.5pep23 determined in "low salt" buffer, all other peptides determined in KPP buffer. For CD scans, data are reported as change in mean residue ellipticity (MRE; units deg cm^2 $dmol^{-1}$, to allow for comparison between peptides of different lengths), as a function of circular dichroism ellipticity over the wavelength range 200 – 300 nm. Blue squares: peptide alone, red circles: equimolar mix of peptide with cJun, grey diamonds: average of cJun and peptide alone, black triangles: cJun alone. **Legend continues overleaf.**

Figure 5.7 continued: CIS display and CIS→PCA selected peptides bind cJun with low affinities. Peptides interact with cJun where MREs for peptide:cJun mixes exceed that of the calculated average trace of peptide and cJun only traces (which represents non-interaction). MRE was calculated using Equation 3 (see 2.9.6.1). cJun traces from Mason *et al.* (2006) are in excellent agreement with that reported here. For CD thermal melts, data are reported as change in MRE as a function of CD ellipticity at 222 nm with temperature. Colours are as for CD scans. Peptides interact with cJun where MRE and/or melting temperature (T_m) for peptide:cJun mixes exceed that of the calculated average trace of peptide and cJun only traces (which represents non-interaction). Heteromeric and homomeric traces were fit to Equations 6 –13 (see 2.9.6.2) to produce data in Table 5.2. For ITC, raw isotherms and fitted data plots are shown, both baseline corrected. On fitted data plots, solid lines represent fits generated by non-linear least-squares fitting to Equation 14 (see 2.9.7). Raw isotherm injections for the buffer into buffer control (blue text label) are translated from the baseline at 0.0 µcal/sec for clarity relative to those for cJun into peptides. **A** 2.5pep1 scans, melts and ITC. **B** 3.5pep22 scans, melts and ITC. **C** 3.5pep23 scans, melts and ITC. **D** 3.4LCIS scans, melts and ITC. **E** 3.4HCIS scans, melts and ITC. **F** 3.4W scans, melts and ITC. 3.4LCIS ITC was unusual, generating two distinct phases to binding that was not typical of dimeric CC formation with cJun ("Phase 1", "Phase 2" and "Phase 2 continued" labelled on ΔH vs. molar ratio plots in **D**).

Table 5.2: CIS display and CIS→PCA selected peptides from the 2.5h, 3.5h and 3.4hFosW libraries interact with cJun with poor affinities. Thermodynamic parameters for peptides in interaction with cJun compared to previous cJun antagonists FosW (CD values from Worrall and Mason, 2011), 4hFosW (values from Crooks *et al.* 2011), and $FosW_{CANDI}$ (values from Mason *et al.* 2007b), and native cFos (values from Worrall and Mason, 2011), as well as CPW, CIS1, CIS21 and LIN20 from Chapters 4 and 3. Data from CD and ITC measurements. θ is raw CD ellipticity (mdeg). Fractional helicity is as calculated using Equation 4 (see 2.9.6.1). TΔS is calculated as ΔH – ΔG from the Gibbs-Helmholtz equation (Equation 11, see 2.9.6.2). CD values are from representative single measurements, typically reproducible in biological replicates to ± 1 °C for T_m, within 5 % for fractional helicity and 222:208 ratio ($\theta_{222}/_{208}$), and within 10 % for K_d and ΔG (data not shown). ITC values are the arithmetic mean of two to three independent titrations ± SDs for FosW, CPW, CIS1 and CIS21, or from single titrations with fitting errors. *Peptides assayed in 'low salt' buffer (10 mM KH_2PO_4/K_2HPO_4, no KF), other peptides assayed in KPP buffer ("low salt"+ 100 mM KF). "ND" = not determined, "NF" = not fit (CD traces lacking lower baseline, ITC fits poor). †Thermal melt traces could not be fit to a 2-state unfolding model. Values given to 2 s.f..

Complex	T_m (°C) (CD)	Peptide–cJun Fractional Helicity (%) (CD)	$K_{d(20°C)}$ (nM) (CD)	$\Delta G_{(20°C)}$ (kcal/mol) (CD)	$K_{d(20°C)}$ (nM) (ITC)	N interaction stoichiometry (ITC)	$\Delta G_{(20°C)}$ (kcal/mol) (ITC)	$\Delta H_{(20°C)}$ (kcal/mol) (ITC)	$T\Delta S_{(20°C)}$ (kcal/mol) (ITC)	Peptide–cJun $\Theta_{222}/_{208}$ (20°C) (CD)
cFos–cJun	16	28	320,000	-5.5	27,000	1.1 ± 0.01	-6.1 ± 0.39	-0.82 ± 0.36	5.3 ± 0.53	0.75
FosW–cJun	63	37	4.0	-11	39 ± 11	0.99 ± 0.08	-9.9 ± 0.16	-10 ± 0.42	-0.46 ± 0.46	1.0
4hFosW–cJun	49	60	490	-8.5	480	1.1 ± 0.01	-8.8 ± 0.10	-14 ± 0.20	-5.3 ± 0.20	0.99
$FosW_{CANDI}$–cJun	52	54	ND	ND	ND	ND	ND	ND	ND	1.0
CPW–cJun*	63	53	0.27	-13	750 ± 270	0.48 ± 0.03	-8.2 ± 0.21	-7.4 ± 0.67	0.81 ± 0.81	0.95
CIS1–cJun*	34	31	14,000	-6.5	25,000 ± 13,000	0.11 ± 0.02	-6.2 ± 0.29	-15 ± 4.4	-9.2 ± 4.6	0.71
CIS21–cJun*	68	48	0.73	-12	770 ± 230	0.52 ± 0.04	-8.2 ± 0.18	-3.6 ± 0.21	4.6 ± 0.37	0.92
LIN20–cJun	25	39	7500	-6.9	55,000 ± 9.7	1.4 ± 0.07	-5.7 ± 0.10	-6.0 ± 0.38	-0.25 ± 0.38	0.81
2.5pep1–cJun*	NF	10	NF	NF	NF	NF	NF	NF	NF	0.47
3.5pep22–cJun*	NF	5.4	NF	NF	NF	NF	NF	NF	NF	0.57
3.5pep23–cJun*	NF	18	NF	NF	ND	ND	ND	ND	ND	0.70
3.4HCIS–cJun	NF	24	NF	NF	ND	ND	ND	ND	ND	0.76
3.4LCIS–cJun	†	38	†	†	NF	NF	NF	NF	NF	0.83
3.4W–cJun	†	44	†	†	54,000 ± 7800	1.8 ± 0.0	-5.8 ± 0.08	-1.4 ± 0.08	4.3 ± 0.12	0.87

CD scans for 2.5pep1 and 3.5pep22 peptides alone unusually were virtually featureless from 200 – 300 nm, though they were not identical to buffer scans (subtracted from all peptide scans presented in this Thesis). As a heteromeric mixture with cJun, 2.5pep1 exhibited weak ellipticity at 222 nm compared to a maximum at ≈200 nm, indicative of a greater proportion of random coil than α-helical structure. Further, the heteromeric trace overlaid the average, indicating non-interaction with cJun and that cJun contributed the majority of the ellipticity to this trace. Peptide 3.5pep22, on the other hand, demonstrated a very weak random coil signature with cJun that was of much lesser ellipticity than the average and so was not typical for non-interaction or interaction, whilst like 2.5pep1 displayed a very low intensity and featureless CD signal alone. Similarly 3.5pep23, which gave a random coil signature alone, gave a heteromeric trace of lesser ellipticity than the average. For 3.4HCIS, the heteromeric mixture with cJun overlaid the average despite reasonable helicity and a 222:208 ratio indicating CC formation for the homodimer (Appendix Table 5.8). Finally, 3.4LCIS and 3.4W demonstrated potential binding to cJun with heteromeric traces fractionally more helical than the average, though 222:208 ratios were close to the ≈0.9 limit expected for CCs (Zhou *et al.*, 1992, Lau *et al.*, 1984). Heteromeric fractional helicity for 3.4LCIS and 3.4W was similar to and slightly higher than FosW respectively, whilst being lower than for $FosW_{CANDI}$ and 4hFosW, but 10 – 16 % higher than cFos–cJun, suggesting more favourable binding of these peptides than cFos.

Heteromeric CD thermal denaturation traces for 2.5pep1, 3.5pep22 and 3.5pep23 overlaid the average indicating non-interaction, in slight contradiction to CD scans for the latter two peptides. The trace for 3.4HCIS with cJun was of lower ellipticity than the average, by a greater margin than in CD scans. Whilst 3.4LCIS and

3.4W melt traces did exceed the average, CD trace line shapes were not typical of those seen for dimeric CCs. In the case of 3.4LCIS, ellipticity of the heteromeric trace increased slightly from -8 °C to ≈20 °C before decreasing (becoming more positive) in the expected sigmoid unfolding curve, with a more pronounced increase in ellipticity over this temperature range for the homomeric sample. As a result, the heteromeric trace did not fit satisfactorily to the expected 2-state unfolding model for CCs. The heteromeric (and homomeric) trace for 3.4W also could not be satisfactorily fit to this model. Thus, thermal melts did not support cJun binding by any of the peptides.

To investigate potential for cJun binding further, ITC was performed on 2.5pep1, 3.5pep22, 3.4LCIS and 3.4W. Titrations for 2.5pep1 and 3.5pep22 indicated very small enthalpies (ΔH) similar to the buffer-to-buffer control, and low signal-to-noise ratios (scattering of data points). For 2.5pep1 this indicated very weak cJun binding as the concentration of 2.5pep1 used (75 μM) would have generated an observably sigmoid binding curve had the K_d been <≈50 μM. For 3.5pep22, the poor enthalpy observed could, on the other hand, be due to the low concentration of 3.5pep22 used (5 μM), which was solubility-limited even in "low salt" buffer (10 mM KH_2PO_4/K_2HPO4). 3.4LCIS curiously generated an isotherm with increasingly enthalpic heat spikes up until a molar ratio of ≈0.5:1 cJun:peptide (3.4LCIS "Phase 1" in Figure 5.7), beyond which titration enthalpies decreased in magnitude as would be expected for binding and increasing saturation of binding sites (3.4LCIS "Phase 2" in Figure 5.7). This second phase continued past a molar ratio of 2.5 as demonstrated by reloading of the ITC syringe with cJun and continuing the titration from a molar ratio of ≈1.4:1 cJun:peptide ("Phase 2 continued" in Figure 5.7). The shape of this isotherm was reproducible (data not shown), and together with CD thermal melts

suggested an alternative binding mode to the expected 2-state model. Finally, 3.4W generated an isotherm with low enthalpy, and a flat line shape despite titration of cJun into 70 µM 3.4W, thus again indicating very weak binding (K_d ≥≈54 µM, which should be treated as approximate when making comparisons). ITC therefore suggested that potentially 2.5pep1, 3.4W and 3.4LCIS may have bound cJun, but so weakly under ITC conditions as to be debatable, or via an unclear binding mode.

The peptides characterised in this Chapter exhibited poor activity when compared to cJun antagonists FosW, $FosW_{CANDI}$, 4hFosW, CPW and CIS21. Finally, LIN20 from Chapter 3 (see 3.2.2.3.4), which at 25 residues was the same length as 3.4HCIS, 3.4LCIS and 3.4W, was able to demonstrate similar/improved affinity for cJun by ITC, despite being a truncated form of FosW without re-randomisation of residues to compensate for a shorter length. Further, the suggested affinity of 3.4W does not exceed that of cFos. Collectively, this indicated that CIS→PCA selection of truncated peptides with re-randomised residues was not successful in identifying a shorter but higher affinity AP-1 antagonist than previously described.

5.3 Discussion

5.3.1 CIS→PCA selected shorter peptides have poor affinities for cJun

It was the aim of this study to use CIS→PCA to select shorter cJun binders than previously derived, with the hypothesis that short peptides with nevertheless high affinity for cJun could be isolated from highly diverse randomised libraries that can be screened by this technique. As discussed in 3.3.6, high affinity in short antagonists necessitates sufficient enthalpy and S_{desolv}, both when adopting a target-binding conformation and binding, to overcome the S_{conf} losses entailed to result in

a favourable binding ΔG. Whether a short peptide could bind cJun with enough affinity to make its smaller size worthwhile was previously unknown, and was tested here with 2.5h and 3.5h/3.4hFosW libraries, where peptides were two or one heptad shorter than FosW, and with even shorter libraries of 12 and 16 residue peptides. However, despite screening ≈6.1 x 10^{10} peptides for each of the five libraries, peptides isolated by CIS display or CIS→PCA demonstrated undesirably low affinities for cJun for the shorter lengths achieved. In isolated peptides, selection of alternative residues to those in corresponding regions of FosW was insufficient to endow improved binding affinity relative to LIN20, the 25mer truncated version of FosW whose residues were not re-randomised. Thus, efforts to derive shorter AP-1 antagonists through library screening were not successful.

5.3.2 Addressing library construction issues

Peptides isolated from the 2.5h and 3.5h libraries in the initial CIS display screening were not all of the expected length due to low frequency but accumulative errors in the ProxiMAX construction process (see 4.3.2). To address this issue, only correct-length peptides identified by sequencing of colonies growing in the PCA Single Step selection were passaged in Competition Selection. This was successful in limiting PCA screened peptides to only those of desired length, whilst being a cheap and facile solution to this issue for the number of colonies observed, though it is noted that for greater numbers of unique peptides selected by CIS display this solution could become labour-intensive. Further, the use of 50 % (for 2.5h) or 75 % (for 3.5h) normal trimethoprim concentrations to encourage correct-length peptide selection requires proper investigation and optimisation in the future to assign

particular affinities of selected peptides (and thus discrete selection stringencies) to the concentration of trimethoprim used, and investigate false positive rates at these stringencies. As an alternative to the 3.5h library, which was considered the library most likely to contain high affinity peptides due to the intermediate length of peptides, the 3.4hFosW library, devoid of ProxiMAX length issues, was constructed and screened. Finally, 12mer and 16mer libraries similarly devoid of length issues allowed screening of some of the shortest peptides reported against cJun to date.

5.3.3 Poor predicted and measured binding of poorly enriched short peptides

Peptide observation frequencies after CIS display indicated poor enrichment generally, except for the 3.4hFosW library and the 3.5h library low stringency selection (see 4.2.4.1). Agadir predicted peptides to be of reasonable helicity generally, with the most helical somewhere between full-length and native cJun binders, to suggest binding could be more favourable than the latter, in contradiction to enrichment analyses. bCIPA did not predict any peptide to have significant affinity for cJun; however, the magnitude of negative T_m values predicted (which can have no physical significance) suggests that bCIPA is inappropriate for use on peptides as short as 18 – 25 amino acids. This is logical given that the training dataset for bCIPA consisted solely of 32mer peptides (Mason *et al.*, 2006), necessitating extrapolations to shorter peptides which no doubt incurs errors. Furthermore, the dataset used did not contain any aromatic residues, and the scoring system of bCIPA itself is limited to considering only those residues previously studied in double mutant cycles (Acharya *et al.*, 2006, Krylov *et al.*, 1998). As a result, bCIPA was inappropriate for predicting the interaction of 3.4HCIS with cJun, and more generally of peptides

containing Ile, Leu, Thr, Val or Ser at supposed ***e*** and ***g*** positions (at outerface positions all residue helicities would be scored), as in 3.5pep22 and 3.5pep23. The BOW algorithm, on the other hand, contains aromatics within its training dataset and considers these residues in stability predictions (Fong *et al.*, 2004). However, BOW could not be used as it has a lower limit for input peptide size of 28 residues, below which predictions are too inaccurate to be useful, due to the use of 34 – 57mer proteins in its training dataset. Sequence alignments to cFos and FosW indicated some similarity to these CC-forming domains, potentially suggesting CC formation by 2.5h, 3.5h and 3.4hFosW selected peptides, though similarities were expected to be higher given that the library design featured amino acids typical of CC domains. Finally, selected peptides were generally suggested to be more aggregation prone than previously derived cJun binding peptides, indicating that secondary selection pressure for non-specific binding/aggregation could well have generally overwhelmed primary selection pressure for cJun affinity in these selections.

In vitro characterisations corroborated poor predicted activities of selected peptides presumably as a result of poor selection efficiency. In the cases of 2.5pep1, 3.5pep22 and 3.5pep23 this could be due to lack of helical structure adoption, in the case of 3.5pep23 by adoption of random coil conformations instead. For 3.5pep22, low concentrations used (limited by poor solubility) could be the source of low CD signal intensity, whilst for 2.5pep1 assayed at 75 µM (a concentration which should have been sufficient for stronger CD signal than that observed) explanations for the lack of CD signal intensity are less forthcoming. In all samples, aggregated precipitates that would result in negligible CD signal compared to pure buffer were not observed upon sample centrifugation, suggesting this was not the source of poor

CD signals. Another possibility is that peptide populations adopted a heterogeneous mixture of structures whose individual CD signals destructively interfered with each other to give an average signal that does not match individual random coil, α-helix, β-turn or β-sheet signals. These suggestions will require further work to investigate. Regardless of the source of poor CD signals, these short peptides did not adopt stable α-helices, as hypothesized could occur upon peptide truncation and loss of helix formation favourability, and thus were unable to supercoil with cJun.

For longer 25mer peptides, despite good helicity and even formation of homodimeric coiled coils, binding was again negligible, indicating that adopting helical structure was not the obstacle preventing binding to cJun, and that instead negligible binding could be the result of low net interaction enthalpy. This again supports the proposition that binding affinity for cJun is spread along the length of the cJun molecule. However, the unusual thermal denaturation/ITC line shapes observed do warrant further investigation for potential binding. Homomeric 3.4HCIS and both homo- and heteromeric 3.4LCIS samples displayed an increase in helicity before cooperative unfolding, which has been observed previously for trimeric coiled coils but was not explained on a structural level (Travisano and Kennan, 2004). The heteromeric thermal melt trace for 3.4W, on the other hand, could be reporting one of a multitude of phenomena, including (but not limited to) two independent transitions from 3-state unfolding of trimeric CCs or from 2-state unfolding of two distinct species. Finally, the 3.4LCIS ITC line shape could indicate formation of higher oligomeric states perhaps involving multiple binding sites (Appendix Figure 5.4). These hypotheses, and others, will of course require future experimental validation.

5.3.4 Selection stringency appropriateness for length of peptides screened

The analyses discussed above indicate that the CIS display selection stringencies applied to the short peptide libraries screened here were likely too high for the cJun-binding affinity of these peptides. Correct-length PCAs making use of lower trimethoprim concentrations to relax stringency and encourage correct-length peptide selection may have helped counter CIS display stringencies being too high, but were ultimately insufficient to allow selection of high affinity short binders. This supports the hypothesis that shorter peptides would have lesser affinity for cJun than longer peptides, such as those of the 3.5h library where the same selection stringencies selected CPW with a K_d of 750 ± 270 nM. The inability to select shorter peptides with comparable affinity to CPW using the same selection stringencies also evidences that binding affinity for cJun is spread along the length of the cJun interface, making isolation of truncated linear antagonists extremely challenging.

5.3.5 Future directions

It is possible that in the future valuable short cJun binders could be selected. Rescreening of the libraries described here at lower selection stringencies (i.e. against more cJun target) could improve selection efficiency and result in successful enrichment of cJun-binding peptides. Whilst selected peptides would likely display lower affinities than originally desired in the screening described here, nevertheless those peptides with affinities at the high end of the affinity range could still be deemed attractive. Such peptides could then be modified, perhaps with helix constraints and/or unnatural amino acids, for example, to develop valuable antagonists from starting peptides of intermediate affinity. By exploring more

thoroughly the affinity ranges of peptides pulled down on different target quantities (as suggested in 4.3.4), more appropriate target quantities for lower affinity short peptides could be ascertained and used for this library re-screening. Alternatively, libraries could be constructed based on 2.5h and 3.5h but with reduced diversity, to allow full coverage of diversity in CIS display screening to isolate rare high affinity binders that were missed in the screening described in this Chapter.

The evidence presented here and in Chapters 3 and 4 collectively suggests that linear unmodified peptides shorter than 32 – 37 amino acids may rarely have sufficient structural stability and enthalpic contacts with cJun to represent attractive AP-1 antagonists. Future work could therefore look to identification of short peptide derivatives by alternative strategies. The successes of library selection systems for high-throughput peptide identification, and of helix constraints for peptide truncation whilst retaining binding affinity, provide the opportunity for combination of these strategies in the form of constrained peptide libraries. These have received growing interest over the past few years as a source of valuable high affinity PPI inhibitors with more attractive pharmacokinetic properties (Heinis and Winter, 2015). Whilst structural constraint of chemically synthesized peptide libraries is perhaps the most flexible embodiment of this approach, constraint of genetically encoded libraries is also possible and has the advantage of being able to screen much higher peptide diversities. For example, libraries of constrained peptides have been produced through *in vitro* post-translational modification of peptides for selection by phage and mRNA display (Heinis *et al.*, 2009, Schlippe *et al.*, 2012). Modification chemistries generally feature use of nucleophilic alkylating agents or reactive oxygen moieties for alkylation of Cysteine side chain sulfurs or condensation reactions with

N-terminal or Lysine amines, to form thioether or amide (or derivative) links in library peptides. Constraint of peptide libraries could ensure short peptides adopt helical structure via entropic pre-organisation to allow the selection process to select optimal residues for cJun binding, as opposed to attempting to simultaneously select residues for both structural stabilisation and binding enthalpy in short linear peptides, potentially two different attributes whose selection could be antagonistic. Modification of the CIS→PCA system described here to achieve constrained peptide libraries would be very attractive based on the aforementioned benefits of CIS→PCA, but would require significant development due to the *in cellulo* nature of peptide selection during the PCA stage. However, CIS display could be modified to incorporate constraints into *in vitro* expressed peptides prior to selection, in much the same way as for phage or mRNA display, to isolate valuable constrained peptides.

In conclusion, shorter linear peptide antagonists against cJun that retained high binding affinity were not identified despite searching of highly diverse libraries by CIS→PCA. In the future, more appropriate selection stringencies could be applied based on the knowledge obtained here of relation of stringencies to peptide affinities, and may reveal rare attractive binders from short peptide libraries. Whilst these might show intermediate affinities alone, they could nevertheless form novel lead compounds for development into more promising AP-1 antagonists through helix constraint, non-natural amino acid substitution and other strategies. Finally, CIS display of constrained peptide libraries could further expedite identification of modified peptides with therapeutic value.

Chapter 6 – Discussion

6.1 Identification and development of novel AP-1 inhibitors

The main focus of this study was to develop novel peptide-based inhibitors of the coiled coil domain of human oncogenic transcription factor Activator Protein-1, a dimer of Fos and Jun proteins known to be the driving agent for tumourigenesis in a wide variety of human tissues. AP-1 activates and represses expression of key proliferation regulatory genes within the nuclei of proliferating cells, making it a challenging target for therapeutic agents to reach. Peptides that displayed high target affinity whilst being small in size to improve envisaged access to nuclear AP-1 were hypothesized to represent desirable progress towards therapeutic agents for the therapy of cancers featuring AP-1 dysregulation. It was therefore the aim of this study to identify and characterise novel peptides that were shorter than previous antagonists and that could sequester the cJun CC domain *in vitro*, to reduce formation of cFos–cJun AP-1 in a therapeutic setting.

To achieve this aim, two approaches were adopted. The first was to modify an existing high affinity but lengthy and so therapeutically unattractive cJun antagonist, known as FosW, using helix constraint modifications. These were hypothesized to entropically pre-organise FosW derivatives for binding cJun and allow truncation whilst retaining binding affinity, which would otherwise be abolished through loss of binding-competent structural propensity. It was also hypothesized that helix constraints might additionally increase protease resistance and cellular uptake of peptides, desirable attributes for the therapeutic end goal. The second was to select alternative truncated peptides from hugely diverse peptide

libraries by novel combination of high throughput *in vitro* CIS display and *in cellulo* PCA peptide library display and selection systems. This aimed to benefit from overlap in the pros and cons of these systems for screening hugely diverse libraries but in an *in vitro* setting, and screening within cell cytoplasms but of greatly reduced library sizes, to select short peptides with attractive affinity for cJun.

Placement of helix constraints in FosW was guided by determination of the crystal structure of FosW–cJun, and constraints were combined with N- and C-terminal truncation. Lactam constraints identified an entropically stabilised peptide **20**, one-third shorter than FosW but retaining approx. 77 % binding free energy. Cellular uptake of this peptide, however, required addition of NLS-Tat delivery moieties which unfortunately reduced AP-1 target gene expression on their own. Replacement of lactam constraints with hydrocarbon constraints generated a peptide which was not taken up into cancer cells, but slightly longer hydrocarbon stapled peptide **18**HC was. This 29mer peptide was 22 % shorter than FosW whilst retaining 88 % of binding free energy (nanomolar affinity) through entropic stabilisation for cJun binding. Finally, this peptide further demonstrated potential AP-1 transcriptional activity inhibition in a breast cancer cell line, though this requires confirmation. Thus, helix constraints allowed shortening of FosW whilst improving cellular uptake and expected protease resistance, to generate an *in cellulo* active peptide with improved therapeutic attractiveness.

Combination of CIS and PCA selection systems was successful in isolating CPW, a novel full-length 32mer antagonist of cJun, which displayed high nanomolar affinity whilst being of a novel sequence to FosW, such that further investigation of

the potential of this peptide is warranted. This demonstrated that an *in vitro* and *in cellulo* system could be combined effectively, and in the future is expected to expedite isolation of valuable peptide ligands against therapeutic targets. For shorter peptide libraries, peptides selected by CIS→PCA unfortunately displayed low affinities for cJun. It was suggested that the selection stringencies applied were too high for the low affinities of library peptides, leading generally to poor enrichment of potentially non-specific peptides. Thus, CIS→PCA demonstrated successful selection of cJun binders from some libraries, and the even greater potential of this system with future refinement of selection stringencies.

Findings from library selection of short peptides, and from lactam constraint and truncation of cJun binders here and from previous study (Rao *et al.*, 2013), suggest that peptide binding enthalpy with cJun is spread along the length of the CC domain, with the consequence that binding enthalpy and free energy are lost with more severe peptide truncations to even more attractive sizes. Helix constraint may provide better structural stabilisation of short peptides than residue re-randomisation and selection based on the greater success of the former approach in generating high affinity short peptides here, but nevertheless the widespread enthalpy along cJun will limit antagonist downsizing that is possible in the future using these strategies. In the future, alternative methods may therefore be required to develop therapeutically attractive peptide antagonists of AP-1 CC domains.

6.2 Issues and approaches adopted to address them

Though CIS→PCA selections of the 3.5h library were successful in identifying high affinity cJun antagonist peptide CPW, issues were encountered with short

peptide library construction using the ProxiMAX approach which lead in particular to DNA constructs encoding longer peptides being screened. Longer peptides were hypothesized to have greater affinity for cJun that correct length peptides, such that desired selection of designed short peptides could suffer, a hypothesis that was supported by preferential selection of longer peptides in CIS display and subsequent PCA. As a result, ProxiMAX as performed here, whilst being advantageous for avoiding peptide library coding redundancy, was suboptimal for libraries of peptides targeting the cJun CC domain, where peptide length and affinity are strongly correlated both due to widespread enthalpy along cJun and the greater structural stability of longer peptides from extended backbone hydrogen bonding. To address these issues, the 3.4hFosW library was constructed by oligonucleotide-directed PCR to replace the ProxiMAX-built 3.5h library and solve the length issues for this library. In addition, an alternative approach was taken to encourage correct length peptide selection from the 2.5h and 3.5h libraries. Sequencing of PCA Single Step selection colonies and transfer only of desired length peptide-containing colonies to Competition Selection was more successful for the 2.5h library which had a lower percentage of longer peptides in, but for both libraries did identify desired length (or slightly shorter for 3.5h) peptides to characterise further. Outstanding issues for further investigation include the lower trimethoprim concentrations used to encourage growth of colonies containing correct length peptides, which could be optimised following derivation of a quantitative relationship between peptide binding affinity and trimethoprim concentration. This could be obtained from *in vitro* characterisation of peptides from efficient selections at different trimethoprim concentrations, by CD or ITC, or via pull-downs on immobilised cJun relative to cJun

antagonists of known affinities. The trimethoprim concentration for 3.5h correct-length peptide selection in particular was suboptimal and could be refined further. This could identify short peptides with affinities lower than originally desired in this study but still high enough to encourage further peptide development. In the future, stringent PAGE purification of ProxiMAX DNA libraries at several stages during construction may limit generation of longer peptides with higher affinity for cJun to improve the utility of ProxiMAX for constructing short peptide libraries.

6.3 Future Directions

In the future, short peptide libraries described here could be rescreened at a lower selection stringency in CIS display, to improve the efficiency of selection of peptides specifically for binding cJun and limit non-specific binder enrichment. This could result in the identification of short peptides with lower affinity than was desired in this study but still high enough to be useful as lead compounds which could be optimised further for higher affinity. Lower selection stringency may also improve the synergy between CIS display and PCA, by limiting CIS display preference for more hydrophobic, non-specific peptides which are subsequently removed by PCA, which may at least partially explain why the low stringency selection here was more successful than the high stringency selection. To improve the overlap between these systems further, the *in vitro* conditions of CIS display could be modified to more accurately mimic *in cellulo* conditions experienced by a therapeutic agent and indeed by peptides transferred to PCA, such as supplementing selections with cell lysate. This may refine the CIS→PCA technique further and allow even more efficient selection of therapeutic-like peptides.

The proposed widespread binding enthalpy along the length of the cJun CC domain may result in future difficulties developing even shorter antagonists due to lack of binding affinity, even where sufficient structure to bind can be retained. To allow further truncation, non-natural amino acid substitution could increase the contribution of enthalpic interactions with cJun, to compensate for loss of enthalpic interactions in truncated regions and so improve binding free energy. This was successful when combined with truncation previously (Rao *et al.*, 2013), and may generate even more promising cJun-binding peptides in the future. Perhaps an even more attractive strategy, based on the proposed limits to antagonist downsizing that could be achieved by helix constraint or residue re-randomisation alone, would be to combine these strategies in a high-throughput form via CIS display selection of constrained peptide libraries. Peptide libraries of natural and/or non-natural amino acids could be constrained through cyclisation reactions performed post-translation, placing constraints in positions identified in this study to be successful, such as at ***b***, ***c***, and ***f*** positions in the heptad repeats of peptides and at peptide termini, with randomisation of amino acids around these constraints, including non-natural amino acids to improve binding enthalpy as discussed above. With careful consideration of the constraint chemistry to ensure compatibility with the CIS display system, constrained and re-randomised peptide libraries could deliver valuable short peptide antagonists of AP-1. Finally, it is worth noting that despite the anticipated benefits of continued shortening of peptide-based cJun antagonists for therapeutic viability, the cellular uptake and expected protease resistance of even 29mer helix constrained peptides demonstrated in this study (i.e. **18**HC) may mean that further antagonist

shortening may not be as crucial for improving therapeutic attractiveness as originally hypothesized.

6.4 Conclusions

In conclusion, this study reports the successful downsizing of previously derived high affinity AP-1 antagonist FosW to a more attractive size for a future anticancer therapeutic goal, whilst retaining the majority of FosW binding free energy and additionally facilitating penetration of breast cancer cells, to potentially inhibit AP-1 *in cellulo* and improve expected protease resistance of this antagonist. This study also reports successful combination of previously described *in vitro* CIS display and *in cellulo* PCA library display and selection systems for isolation of peptide ligands against therapeutic protein targets, including the selection of a novel high affinity AP-1 antagonist worthy of future exploration. Peptides identified and developed by these two approaches are anticipated with further optimisation to form the basis of therapeutic agents for the treatment of cancers featuring AP-1 dysregulation. Finally, the described successful combination of CIS display and PCA, and the proposed future refinements of this technique, could open the door for combination of other library selection systems and their optimisation, to facilitate novel antagonist peptide selections under conditions that better represent the *in vivo* situations encountered by therapeutic agents. Such advances could expedite the development of therapeutic molecules for therapy of numerous human diseases.

References

Acharya, A., Rishi, V. & Vinson, C. (2006). Stability of 100 homo and heterotypic coiled-coil ***a–a'*** pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry,* 45**,** 11324-11332.

Agrawal, P., Bhalla, S., Usmani, S. S., Singh, S., Chaudhary, K., Raghava, G. P. S. & Gautam, A. (2016). CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Research,* 44**,** D1098-D1103.

Aikawa, Y., Morimoto, K., Yamamoto, T., Chaki, H., Hashiramoto, A., Narita, H., Hirono, S. & Shiozawa, S. (2008). Treatment of arthritis with a selective inhibitor of c-Fos/activator protein-1. *Nature Biotechnology,* 26**,** 817-823.

Albanese, C., D'Amico, M., Reutens, A. T., Fu, M. F., Watanabe, G., Lee, R. J., Kitsis, R. N., Henglein, B., Avantaggiati, M., Somasundaram, K., Thimmapaya, B. & Pestell, R. G. (1999). Activation of the cyclin D1 gene by the EPA-associated protein p300 through AP-1 inhibits cellular apoptosis. *Journal of Biological Chemistry,* 274**,** 34186-34195.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology,* 215**,** 403-410.

Andrew, C. D., Bhattacharjee, S., Kokkoni, N., Hirst, J. D., Jones, G. R. & Doig, A. J. (2002). Stabilizing interactions between aromatic and basic side chains in alpha-helical peptides and proteins. Tyrosine effects on helix circular dichroism. *Journal of the American Chemical Society,* 124**,** 12706-12714.

Andrew, C. D., Penel, S., Jones, G. R. & Doig, A. J. (2001). Stabilizing nonpolar/polar side-chain interactions in the alpha-helix. *Proteins: Structure Function and Genetics,* 45**,** 449-455.

Andrews, S. 2016. *FastQC* [Online]. Babraham Institute. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, Accessed 23/05/16.

Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P. & Karin, M. (1987). Phorbol ester inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell,* 49**,** 729-739.

Angel, P. & Karin, M. (1991). The role of Jun, Fos, and the AP-1 complex in cell-proliferation and transformation. *Biochimica et Biophysica Acta,* 1072**,** 129-157.

Arap, W., Kolonin, M. G., Trepel, M., Lahdenranta, J., Cardo-Vila, M., Giordano, R. J., Mintz, P. J., Ardelt, P. U., Yao, V. J., Vidal, C. I., Chen, L., Flamm, A., Valtanen, H., Weavind, L. M., Hicks, M. E.*, et al.* (2002). Steps toward mapping the human vasculature by phage display. *Nature Medicine,* 8**,** 121-127.

Arias, J., Alberts, A. S., Brindle, P., Claret, F. X., Smeal, T., Karin, M., Feramisco, J. & Montminy, M. (1994). Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. *Nature,* 370**,** 226-229.

Ashraf, M., Frigotto, L., Smith, M. E., Patel, S., Hughes, M. D., Poole, A. J., Hebaishi, H. R. M., Ullman, C. G. & Hine, A. V. (2013). ProxiMAX randomization: a new technology for non-degenerate saturation mutagenesis of contiguous codons. *Biochemical Society Transactions,* 41**,** 1189-U250.

Bains, N. P. S., Wilce, J. A., Heuer, K. H., Tunstall, M., Mackay, J. P., Bennett, M. R., Weiss, A. S. & King, G. F. (1997). Zipping up transcription factors: Rational design of anti-Jun and anti-Fos peptides. *Letters in Peptide Science,* 4**,** 67-77.

Bakin, A. V. & Curran, T. (1999). Role of DNA 5-methylcytosine transferase in cell transformation by fos. *Science,* 283**,** 387-390.

Baxter, D., Ullman, C. G. & Mason, J. M. (2014). Library construction, selection and modification strategies to generate therapeutic peptide-based modulators of protein-protein interactions. *Future Medicinal Chemistry,* 6**,** 2073-2092.

Benkirane, N., Friede, M., Guichard, G., Briand, J. P., Vanregenmortel, M. H. V. & Muller, S. (1993). Antigenicity and immunogenicity of modified synthetic peptides containing D-amino acid residues - antibodies to a D-enantiomer do recognize the parent L-hexapeptide and reciprocally. *Journal of Biological Chemistry,* 268**,** 26279-26285.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, R. K., Cox, A. J.*, et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-59.

Bernal, F., Tyler, A. F., Korsmeyer, S. J., Walensky, L. D. & Verdine, G. L. (2007). Reactivation of the p53 tumor suppressor pathway by a stapled p53 peptide. *Journal of the American Chemical Society,* 129**,** 2456-2457.

Bird, G. H., Gavathiotis, E., LaBelle, J. L., Katz, S. G. & Walensky, L. D. (2014). Distinct BimBH3 (BimSAHB) Stapled Peptides for Structural and Cellular Studies. *ACS Chemical Biology,* 9**,** 831-837.

Bissantz, C., Kuhn, B. & Stahl, M. (2010). A Medicinal Chemist's Guide to Molecular Interactions. *Journal of Medicinal Chemistry,* 53**,** 5061-5084.

Brown, P. H., Chen, T. K. & Birrer, M. J. (1994). Mechanism of action of a dominant-negative mutant of c-Jun. *Oncogene,* 9**,** 791-799.

Burkhard, P., Ivaninskii, S. & Lustig, A. (2002). Improving coiled-coil stability by optimizing ionic interactions. *Journal of Molecular Biology,* 318**,** 901-910.

Campbell, K. M. & Lumb, K. J. (2002). Complementation of buried lysine and surface polar residues in a designed heterodimeric coiled coil. *Biochemistry,* 41**,** 7169-7175.

Chakrabartty, A., Doig, A. J. & Baldwin, R. L. (1993a). Helix capping propensities in peptides parallel those in proteins. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 11332-11336.

Chakrabartty, A., Kortemme, T., Padmanabhan, S. & Baldwin, R. L. (1993b). Aromatic side-chain contribution to far-ultraviolet circular-dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry,* 32**,** 5560-5565.

Chang, Y. S., Graves, B., Guerlavais, V., Tovar, C., Packman, K., To, K. H., Olson, K. A., Kesavan, K., Gangurde, P., Mukherjee, A., Baker, T., Darlak, K., Elkin, C., Filipovic, Z., Qureshi, F. Z.*, et al.* (2013). Stapled alpha-helical peptide drug development: A potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** E3445-E3454.

Chapman, R. N., Dimartino, G. & Arora, P. S. (2004). A highly stable short alpha-helix constrained by a main-chain hydrogen-bond surrogate. *Journal of the American Chemical Society,* 126**,** 12252-12253.

Chen, L., Glover, J. N. M., Hogan, P. G., Rao, A. & Harrison, S. C. (1998). Structure of the DNA binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature,* 392**,** 42-48.

Chen, T. K., Smith, L. M., Gebhardt, D. K., Birrer, M. J. & Brown, P. H. (1996). Activation and inhibition of the AP-1 complex in human breast cancer cells. *Molecular Carcinogenesis,* 15**,** 215-226.

Chiu, R., Angel, P. & Karin, M. (1989). Jun-B differs in its biological properties from, and is a negative regulator of, c-Jun. *Cell,* 59**,** 979-986.

Chiu, R., Boyle, W. J., Meek, J., Smeal, T., Hunter, T. & Karin, M. (1988). The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. *Cell,* 54**,** 541-552.

Chu, Q., Moellering, R. E., Hilinski, G. J., Kim, Y. W., Grossmann, T. N., Yeh, J. T. H. & Verdine, G. L. (2015). Towards understanding cell penetration by stapled peptides. *Medchemcomm,* 6**,** 111-119.

Corbi-Verge, C. & Kim, P. M. (2016). Motif mediated protein-protein interactions as drug targets. *Cell Communication and Signaling,* 14**,** 12.

Craik, D. J., Fairlie, D. P., Liras, S. & Price, D. (2013). The Future of Peptide-based Drugs. *Chemical Biology & Drug Design,* 81**,** 136-147.

Crick, F. H. C. (1953). The Packing of α-Helices: Simple Coiled-Coils. *Acta Crystallographica,* 6**,** 689-697.

Crooks, R. O. 2013. *Designing inhibitors of activator protein 1 based on the leucine zipper motif.* PhD Thesis, University of Essex.

Crooks, R. O., Rao, T. & Mason, J. M. (2011). Truncation, randomization, and selection: generation of a reduced length c-Jun antagonist that retains high interaction stability. *Journal of Biological Chemistry,* 286**,** 29470-9.

de Araujo, A. D., Hoang, H. N., Kok, W. M., Diness, F., Gupta, P., Hill, T. A., Driver, R. W., Price, D. A., Liras, S. & Fairlie, D. P. (2014). Comparative alpha-Helicity of Cyclic Pentapeptides in Water. *Angewandte Chemie International Edition,* 53**,** 6965-6969.

De Crescenzo, G., Litowski, J. R., Hodges, R. S. & O'Connor-McCourt, M. D. (2003). Real-time monitoring of the interactions of two-stranded de novo designed coiled-coils: Effect of chain length on the kinetic and thermodynamic constants of binding. *Biochemistry,* 42**,** 1754-1763.

Debaisieux, S., Rayne, F., Yezid, H. & Beaumelle, B. (2012). The Ins and Outs of HIV-1 Tat. *Traffic,* 13**,** 355-363.

Denault, M. & Pelletier, J. N. 2007. Protein library design and screening: Working out the probabilities. *Methods in Molecular Biology.* Clifton, N.J.: Humana Press.

Deng, T. L. & Karin, M. (1993). JunB differs from c-Jun in its DNA-binding and dimerization domains, and represses c-Jun by formation of inactive heterodimers. *Genes & Development,* 7**,** 479-490.

Dietz, G. P. H. & Bahr, M. (2004). Delivery of bioactive molecules into the cell: The Trojan horse approach. *Molecular and Cellular Neuroscience,* 27**,** 85-131.

Doig, A. J. & Baldwin, R. L. (1995). N- and C- capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Science,* 4**,** 1325-1336.

Dragan, A. I., Potekhin, S. A., Sivolob, A., Lu, M. & Privalov, P. L. (2004). Kinetics and thermodynamics of the unfolding and refolding of the three-stranded alpha-helical coiled coil, Lpp-56. *Biochemistry,* 43**,** 14891-14900.

Dragan, A. I. & Privalov, P. L. (2002). Unfolding of a leucine zipper is not a simple two-state transition. *Journal of Molecular Biology,* 321**,** 891-908.

Du, H., Fuh, R. C. A., Li, J. Z., Corkan, L. A. & Lindsey, J. S. (1998). PhotochemCAD: A computer-aided design and research tool in photochemistry. *Photochemistry and Photobiology,* 68**,** 141-142.

Duchardt, F., Fotin-Mleczek, M., Schwarz, H., Fischer, R. & Brock, R. (2007). A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. *Traffic,* 8**,** 848-866.

Edwards, A. L., Meijer, D. H., Guerra, R. M., Molenaar, R. J., Alberta, J. A., Bernal, F., Bird, G. H., Stiles, C. D. & Walensky, L. D. (2016). Challenges in Targeting a Basic Helix-Loop-Helix Transcription Factor with Hydrocarbon-Stapled Peptides. *ACS Chemical Biology,* 11**,** 3146-3153.

Eferl, R. & Wagner, E. F. (2003). AP-1: a double-edged sword in tumorigenesis. *Nature Reviews: Cancer,* 3**,** 859-68.

Eldridge, B., Cooley, R. N., Odegrip, R., McGregor, D. P., FitzGerald, K. J. & Ullman, C. G. (2009). An in vitro selection strategy for conferring protease resistance to ligand binding peptides. *Protein Engineering, Design & Selection,* 22**,** 691-698.

Ferrara, P., Andermarcher, E., Bossis, G., Acquaviva, C., Brockly, F., Jariel-Encontre, I. & Piechaczyk, M. (2003). The structural determinants responsible for c-Fos protein proteasomal degradation differ according to the conditions of expression. *Oncogene,* 22**,** 1461-1474.

Fong, J. H., Keating, A. E. & Singh, M. (2004). Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biology,* 5**,** 10.

Fosgerau, K. & Hoffmann, T. (2015). Peptide therapeutics: current status and future directions. *Drug Discovery Today,* 20**,** 122-128.

Frank, S., Lustig, A., Schulthess, T., Engel, J. & Kammerer, R. A. (2000). A distinct seven-residue trigger sequence is indispensable for proper coiled-coil formation of the human macrophage scavenger receptor oligomerization domain. *Journal of Biological Chemistry,* 275**,** 11672-11677.

Franklin, C. C., McCulloch, A. V. & Kraft, A. S. (1995). In-vitro association between the Jun protein family and the general transcription factors, TBP and TFIIB. *Biochemical Journal,* 305**,** 967-974.

Frigotto, L., Smith, M. E., Brankin, C., Sedani, A., Cooper, S. E., Kanwar, N., Evans, D., Svobodova, S., Baar, C., Glanville, J., Ullman, C. G. & Hine, A. V. (2015). Codon-Precise, Synthetic, Antibody Fragment Libraries Built Using Automated Hexamer Codon Additions and Validated through Next Generation Sequencing. *Antibodies,* 4**,** 88-102.

Fujii, Y., Shimizu, T., Toda, T., Yanagida, M. & Hakoshima, T. (2000). Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nature Structural Biology,* 7**,** 889-893.

Glover, J. N. M. & Harrison, S. C. (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. *Nature,* 373**,** 257-261.

Gonzalez, L., Woolfson, D. N. & Alber, T. (1996). Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nature Structural Biology,* 3**,** 1011-1018.

Goodman, M. & Chorev, M. (1979). Concept of linear modified retro-peptide structures. *Accounts of Chemical Research,* 12**,** 1-7.

Graddis, T. J., Myszka, D. G. & Chaiken, I. M. (1993). Controlled formation of model homodimer and heterodimer coiled-coil polypeptides. *Biochemistry,* 32**,** 12664-12671.

Grigoryan, G. 2016. *DrawCoil 1.0* [Online]. Dartmouth College. Available: http://www.grigoryanlab.org/drawcoil/, Accessed 23/05/16.

Grigoryan, G., Reinke, A. W. & Keating, A. E. (2009). Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature,* 458**,** 859-U2.

Grigoryev, Y. (2013). Stapled peptide to enter human testing, but affinity questions remain. *Nature Medicine,* 19**,** 120-120.

Hadley, E. B. & Gellman, S. H. (2006). An antiparallel alpha-helical coiled-coil model system for rapid assessment of side-chain recognition at the hydrophobic interface. *Journal of the American Chemical Society,* 128**,** 16444-16445.

Harbury, P. B., Kim, P. S. & Alber, T. (1994). Crystal structure of an isoleucine-zipper trimer. *Nature,* 371**,** 80-83.

Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993). A switch between 2-stranded, 3-stranded and 4-stranded coiled coils in GCN4 leucine-zipper mutants. *Science,* 262**,** 1401-1407.

Harrison, R. S., Shepherd, N. E., Hoang, H. N., Ruiz-Gomez, G., Hill, T. A., Driver, R. W., Desai, V. S., Young, P. R., Abbenante, G. & Fairlie, D. P. (2010). Downsizing human, bacterial, and viral proteins to short water-stable alpha helices that maintain biological potency. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 11686-11691.

Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nature Structural Biology,* 10**,** 45-52.

Heinis, C., Rutherford, T., Freund, S. & Winter, G. (2009). Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nature Chemical Biology,* 5**,** 502-507.

Heinis, C. & Winter, G. (2015). Encoded libraries of chemically modified peptides. *Current Opinion in Chemical Biology,* 26**,** 89-98.

Herce, H. D., Garcia, A. E. & Cardoso, M. C. (2014). Fundamental Molecular Mechanism for the Cellular Uptake of Guanidinium-Rich Molecules. *Journal of the American Chemical Society,* 136**,** 17459-17467.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene,* 77**,** 51-59.

Hoang, H. N., Driver, R. W., Beyer, R. L., Hill, T. A., D de Araujo, A., Plisson, F., Harrison, R. S., Goedecke, L., Shepherd, N. E. & Fairlie, D. P. (2016). Helix Nucleation by the Smallest Known alpha-Helix in Water. *Angewandte Chemie International Edition,* 55**,** 8275-9.

Hsieh, T. H., Hsu, C. Y., Tsai, C. F., Chiu, C. C., Liang, S. S., Wang, T. N., Kuo, P. L., Long, C. Y. & Tsai, E. M. (2016). A novel cell-penetrating peptide suppresses breast

tumorigenesis by inhibiting beta-catenin/LEF-1 signaling. *Scientific Reports,* 6**,** 12.

Jalali-Yazdi, F., Lai, L. H., Takahashi, T. T. & Roberts, R. W. (2016). High-Throughput Measurement of Binding Kinetics by mRNA Display and Next-Generation Sequencing. *Angewandte Chemie-International Edition,* 55**,** 4007-4010.

Jin, X., Song, X., Li, L., Wang, Z., Tao, Y., Deng, L., Tang, M., Yi, W. & Cao, Y. (2007). Blockade of AP-1 activity by dominant-negative TAM67 can abrogate the oncogenic phenotype in latent membrane protein 1-positive human nasopharyngeal carcinoma. *Molecular Carcinogenesis,* 46**,** 901-911.

Jochim, A. L. & Arora, P. S. (2010). Systematic Analysis of Helical Protein Interfaces Reveals Targets for Synthetic Inhibitors. *ACS Chemical Biology,* 5**,** 919-923.

Johnson, L. M. & Gellman, S. H. 2013. alpha-Helix Mimicry with alpha/beta-Peptides. *In:* Keating, A. E. (ed.) *Methods in Protein Design.* San Diego: Elsevier Academic Press Inc.

Kaplan, J. B., Reinke, A. W. & Keating, A. E. (2014). Increasing the affinity of selective bZIP-binding peptides through surface residue redesign. *Protein Science,* 23**,** 940-953.

Kaspar, A. A. & Reichert, J. M. (2013). Future directions for peptide therapeutics development. *Drug Discovery Today,* 18**,** 807-817.

Kawakami, T. & Murakami, H. (2012). Genetically encoded libraries of nonstandard peptides. *Journal of Nucleic Acids,* 2012.

Ke, S. H. & Madison, E. L. (1997). Rapid and efficient site-directed mutagenesis by single-tube 'megaprimer' PCR method. *Nucleic Acids Research,* 25**,** 3371-3372.

Kelso, M. J., Beyer, R. L., Hoangt, H. N., Lakdawala, A. S., Snyder, J. P., Oliver, W. V., Robertson, T. A., Appleton, T. G. & Fairlie, D. P. (2004). alpha-turn mimetics: short peptide alpha-Helices composed of cyclic metallopentapeptide modules. *Journal of the American Chemical Society,* 126**,** 4828-4842.

Kentsis, A. & Sosnick, T. R. (1998). Trifluoroethanol promotes helix formation by destabilizing backbone exposure: Desolvation rather than native hydrogen bonding defines the kinetic pathway of dimeric coiled coil folding. *Biochemistry,* 37**,** 14613-14622.

Kerppola, T. K. (2006). Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nature Protocols,* 1**,** 1278-1286.

Kilby, J. M., Lalezari, J. P., Eron, J. J., Carlson, M., Cohen, C., Arduino, R. C., Goodgame, J. C., Gallant, J. E., Volberding, P., Murphy, R. L., Valentine, F., Saag, M. S., Nelson, E. L., Sista, P. R. & Dusek, A. (2002). The safety, plasma pharmacokinetics, and antiviral activity of subcutaneous enfuvirtide (T-20), a peptide inhibitor of gp41-mediated virus fusion, in HIV-infected adults. *AIDS Research and Human Retroviruses,* 18**,** 685-693.

Kim, Y. W., Grossmann, T. N. & Verdine, G. L. (2011). Synthesis of all-hydrocarbon stapled alpha-helical peptides by ring-closing olefin metathesis. *Nature Protocols,* 6**,** 761-771.

Kohler, J. J. & Schepartz, A. (2001). Kinetic studies of Fos·Jun·DNA complex formation: DNA binding prior to dimerization. *Biochemistry,* 40**,** 130-142.

Kohn, W. D., Kay, C. M. & Hodges, R. S. (1997). Salt effects on protein stability: Two-stranded alpha-helical coiled-coils containing inter- or intrahelical ion pairs. *Journal of Molecular Biology,* 267**,** 1039-1052.

Kouzarides, T. & Ziff, E. (1988). The role of the Leucine Zipper in the Fos Jun interaction. *Nature,* 336**,** 646-651.

Krylov, D., Barchi, J. & Vinson, C. (1998). Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *Journal of Molecular Biology,* 279**,** 959-972.

Lacroix, E., Viguera, A. R. & Serrano, L. (1998). Elucidating the folding problem of alpha-helices: Local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *Journal of Molecular Biology,* 284**,** 173-191.

Laehnemann, D., Borkhardt, A. & McHardy, A. C. (2016). Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Briefings in Bioinformatics,* 17**,** 154-179.

Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. (1988a). Isolation of a recombinant copy of the gene encoding C/EBP. *Genes & Development,* 2**,** 786-800.

Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988b). The leucine zipper - a hypothetical structure common to a new class of DNA-binding proteins. *Science,* 240**,** 1759-1764.

Lanford, R. E., Kanda, P. & Kennedy, R. C. (1986). Induction of nuclear transport with a synthetic peptide homologous to the SV40 T-antigen transport signal. *Cell,* 46**,** 575-582.

Lanning, M. & Fletcher, S. (2013). Recapitulating the alpha-helix: nonpeptidic, low-molecular-weight ligands for the modulation of helix-mediated protein-protein interactions. *Future Medicinal Chemistry,* 5**,** 2157-2174.

Lau, S. Y. M., Taneja, A. K. & Hodges, R. S. (1984). Synthesis of a model protein of defined secondary and quaternary structure - effect of chain length on the stabilization and formation of 2-stranded alpha-helical coiled coils. *Journal of Biological Chemistry,* 259**,** 3253-3261.

Leaner, V. D., Chick, J. F., Donninger, H., Linniola, I., Mendoza, A., Khanna, C. & Birrer, M. J. (2009). Inhibition of AP-1 Transcriptional Activity Blocks the Migration, Invasion, and Experimental Metastasis of Murine Osteosarcoma. *American Journal of Pathology,* 174**,** 265-275.

Li, Y. C., Rodewald, L. W., Hoppmann, C., Wong, E. T., Lebreton, S., Safar, P., Patek, M., Wang, L., Wertman, K. F. & Wahl, G. M. (2014). A Versatile Platform to Analyze Low-Affinity and Transient Protein-Protein Interactions in Living Cells in Real Time. *Cell Reports,* 9**,** 1946-1958.

Lifson, S. & Roig, A. (1961). On the theory of helix-coil transitions in polypeptides. *Journal of Chemical Physics,* 34**,** 1963-1974.

Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews,* 23**,** 3-25.

Liskamp, R. M. J., Rijkers, D. T. S., Kruijtzer, J. A. W. & Kemmink, J. (2011). Peptides and Proteins as a Continuing Exciting Source of Inspiration for Peptidomimetics. *ChemBioChem,* 12**,** 1626-1653.

Liu, Y. M., Lu, C. H., Shen, Q., Munoz-Medellin, D., Kim, H. & Brown, P. H. (2004). AP-1 blockade in breast cancer cells causes cell cycle arrest by suppressing G1 cyclin expression and reducing cyclin-dependent kinase activity. *Oncogene,* 23**,** 8238-8246.

Liu, Y. M., Ludes-Meyers, J., Zhang, Y., Munoz-Medellin, D., Kim, H. T., Lu, C. H., Ge, G. Q., Schiff, R., Hilsenbeck, S. G., Osborne, C. K. & Brown, P. H. (2002). Inhibition of AP-1 transcription factor causes blockade of multiple signal transduction pathways and inhibits breast cancer growth. *Oncogene,* 21**,** 7680-7689.

Lloyd, A., Yancheva, N. & Wasylyk, B. (1991). Transformation suppressor activity of a Jun transcription factor lacking its activation domain. *Nature,* 352**,** 635-8.

Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology,* 285**,** 2177-2198.

Lopez-Bergami, P., Lau, E. & Ronai, Z. (2010). Emerging roles of ATF2 and the dynamic AP1 network in cancer. *Nature Reviews Cancer,* 10**,** 65-76.

Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., Degrado, W. F. & Eisenberg, D. (1993). Crystal structure of a synthetic triple-stranded alpha-helical bundle. *Science,* 259**,** 1288-1293.

Lumb, K. J. & Kim, P. S. (1995). A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled-coil. *Biochemistry,* 34**,** 8642-8648.

Luo, P. Z. & Baldwin, R. L. (1997). Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry,* 36**,** 8413-8421.

Lupas, A. N. & Gruber, M. (2005). The structure of alpha-helical coiled coils. *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers,* 70**,** 37-+.

Madala, P. K., Tyndall, J. D. A., Nall, T. & Fairlie, D. P. (2010). Update 1 of: Proteases Universally Recognize Beta Strands In Their Active Sites. *Chemical Reviews,* 110**,** PR1-PR31.

Makhatadze, G. I. 2005. Thermodynamics of alpha-helix formation. *In:* Baldwin, R. L. & Baker, D. (eds.) *Peptide Solvation and H-Bonds.* San Diego: Elsevier Academic Press Inc.

Malnou, C. E., Salem, T., Brockly, F., Wodrich, H., Piechaczyk, M. & Jariel-Encontre, I. (2007). Heterodimerization with jun family members regulates c-fos nucleocytoplasmic traffic. *Journal of Biological Chemistry,* 282**,** 31046-31059.

Maritz, M. F., van der Watt, P. J., Holderness, N., Birrer, M. J. & Leaner, V. D. (2011). Inhibition of AP-1 suppresses cervical cancer cell proliferation and is associated with p21 expression. *Biological Chemistry,* 392**,** 439-448.

Masai, H., Kaziro, Y. & Arai, K. I. (1983). Definition of OriR, the minimum DNA segment essential for initiation of R1 plasmid replication in vitro. *Proceedings of the National Academy of Sciences of the United States of America,* 80**,** 6814-6818.

Mason, J. M. & Arndt, K. M. (2004). Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem,* 5**,** 170-6.

Mason, J. M., Hagemann, U. B. & Arndt, K. M. (2007a). Improved stability of the Jun-Fos Activator Protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis. *Journal of Biological Chemistry,* 282**,** 23015-24.

Mason, J. M., Mueller, K. M. & Arndt, K. M. (2007b). Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. *Biochemistry,* 46**,** 4804-4814.

Mason, J. M., Schmitz, M. A., Muller, K. & Arndt, K. M. (2006). Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 8989-8994.

McClain, D. L., Gurnon, D. G. & Oakley, M. G. (2002). Importance of potential interhelical salt-bridges involving interior residues for coiled-coil stability and quaternary structure. *Journal of Molecular Biology,* 324**,** 257-270.

Miller, A. D., Curran, T. & Verma, I. M. (1984). c-Fos protein can induce cellular transformation - a novel mechanism of activation of a cellular oncogene. *Cell,* 36**,** 51-60.

Monera, O. D., Kay, C. M. & Hodges, R. S. (1994). Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in 2-stranded alpha-helical coiled-coils. *Biochemistry,* 33**,** 3862-3871.

Monera, O. D., Zhou, N. E., Lavigne, P., Kay, C. M. & Hodges, R. S. (1996). Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *Journal of Biological Chemistry,* 271**,** 3995-4001.

Nieuwlandt, D. (2000). In vitro selection of functional nucleic acid sequences. *Current Issues in Molecular Biology,* 2**,** 9-16.

Novagen. 2016a. *TB045VM pET-15b Vector Map* [Online]. Nottingham, UK. Available: http://www.merckmillipore.com/GB/en/product/pET-15b-DNA---Novagen,EMD_BIO-69661#documentation, Accessed 15/07/16.

Novagen. 2016b. *TB074VM pET-28a-c(+) Vector Map* [Online]. Nottingham, UK. Available: https://www.merckmillipore.com/GB/en/product/pET-28a%28%2B%29-DNA---Novagen,EMD_BIO-69864?bd=1#documentation, Accessed 15/07/16.

Novagen. 2016c. *TB147VM pET-33b(+) Vector Map* [Online]. Nottingham, UK. Available: http://www.merckmillipore.com/INTERSHOP/web/WFS/Merck-GB-Site/en_US/-/USD/ViewParametricSearch-SimpleOfferSearch?SearchTerm=pet-33&SelectedSearchResult=SFDocumentSearch&SearchContextPageletUUID= , Accessed 15/07/16.

O'Neil, K. T. & Degrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science,* 250**,** 646-651.

O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a 2-stranded, parallel coiled coil. *Science,* 254**,** 539-544.

O'Shea, E. K., Rutkowski, R., Stafford, W. F. & Kim, P. S. (1989). Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. *Science,* 245**,** 646-648.

Oakley, M. G. & Kim, P. S. (1998). A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry,* 37**,** 12603-12610.

Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlman, P. A., Ullman, C., FitzGerald, K. & McGregor, D. (2004). CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 2806-2810.

Olive, M., Krylov, D., Echlin, D. R., Gardner, K., Taparowsky, E. & Vinson, C. (1997). A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. *Journal of Biological Chemistry,* 272**,** 18586-18594.

Padmanabhan, S., York, E. J., Stewart, J. M. & Baldwin, R. L. (1996). Helix propensities of basic amino acids increase with the length of the side-chain. *Journal of Molecular Biology,* 257**,** 726-734.

Park, S., Lee, D. K., Whang, Y. H. & Yang, C. H. (2000). Momordin I, a compound of ampelopsis radix, inhibits AP-1 activation induced by phorbol ester. *Cancer Letters,* 152**,** 1-8.

Pauling, L., Corey, R. B. & Branson, H. R. (1951). The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America,* 37**,** 205-11.

Pawar, A. P., DuBay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M. & Dobson, C. M. (2005). Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *Journal of Molecular Biology,* 350**,** 379-392.

Pazos, E., Portela, C., Penas, C., Vazquez, M. E. & Mascarenas, J. L. (2015). Peptide-DNA conjugates as tailored bivalent binders of the oncoprotein c-Jun. *Organic & Biomolecular Chemistry,* 13**,** 5385-5390.

Pelay-Gimeno, M., Glas, A., Koch, O. & Grossmann, T. N. (2015). Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angewandte Chemie-International Edition,* 54**,** 8896-8927.

Pelletier, J. N., Arndt, K. M., Pluckthun, A. & Michnick, S. W. (1999). An in vivo library-versus-library selection of optimized protein-protein interactions. *Nature Biotechnology,* 17**,** 683-690.

Potapov, V., Kaplan, J. B. & Keating, A. E. (2015). Data-Driven Prediction and Design of bZIP Coiled-Coil Interactions. *PLoS Computational Biology,* 11**,** 28.

Potocky, T. B., Menon, A. K. & Gellman, S. H. (2003). Cytoplasmic and nuclear delivery of a TAT-derived peptide and a beta-peptide after endocytic uptake into HeLa cells. *Journal of Biological Chemistry,* 278**,** 50188-50194.

Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Advances in Protein Chemistry,* 33**,** 167-241.

Qi, Y. Z. & Chilkoti, A. (2015). Protein-polymer conjugation - moving beyond PEGylation. *Current Opinion in Chemical Biology,* 28**,** 181-193.

Rao, T., Ruiz-Gómez, G., Hill, T. A., Hoang, H. N., Fairlie, D. P. & Mason, J. M. (2013). Truncated and helix-constrained peptides with high affinity and specificity for the cFos coiled-coil of AP-1. *PloS One,* 8**,** e59415-e59426.

Remy, I. & Michnick, S. W. (1999). Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 5394-5399.

Robertson, N. S. & Jamieson, A. G. (2015). Regulation of protein–protein interactions using stapled peptides. *Reports in Organic Chemistry,* 2015**,** 65-74.

Rohl, C. A. & Baldwin, R. L. (1997). Comparison of NH exchange and circular dichroism as techniques for measuring the parameters of the helix-coil transition in peptides. *Biochemistry,* 36**,** 8435-8442.

Rohl, C. A., Scholtz, J. M., York, E. J., Stewart, J. M. & Baldwin, R. L. (1992). Kinetics of amide proton-exchange in helical peptides of varying chain lengths - interpretation by the Lifson-Roig equation. *Biochemistry,* 31**,** 1263-1269.

Ryseck, R. P. & Bravo, R. (1991). c-Jun, Jun-B, and Jun-D differ in their binding affinities to AP-1 and CRE consensus sequences - effect of Fos proteins. *Oncogene,* 6**,** 533-542.

Sassone-Corsi, P., Ransone, L. J., Lamph, W. W. & Verma, I. M. (1988). Direct interaction between Fos and Jun nuclear oncoproteins - role of the Leucine Zipper domain. *Nature,* 336**,** 692-695.

Sato, A. K., Viswanathan, M., Kent, R. B. & Wood, C. R. (2006). Therapeutic peptides: technological advances driving peptides into development. *Current Opinion in Biotechnology,* 17**,** 638-642.

Schafmeister, C. E., Po, J. & Verdine, G. L. (2000). An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides. *Journal of the American Chemical Society,* 122**,** 5891-5892.

Schlippe, Y. V. G., Hartman, M. C. T., Josephson, K. & Szostak, J. W. (2012). In Vitro Selection of Highly Modified Cyclic Peptides That Act as Tight Binding Inhibitors. *Journal of the American Chemical Society,* 134**,** 10469-10477.

Schrodinger, L. L. C. (2013). The PyMOL Molecular Graphics System. Open Source Version 1.3 ed.

Schuermann, M., Neuberg, M., Hunter, J. B., Jenuwein, T., Ryseck, R. P., Bravo, R. & Muller, R. (1989). The leucine repeat motif in Fos protein mediates complex formation with Jun AP-1 and is required for transformation. *Cell,* 56**,** 507-516.

Schumacher, T. N. M., Mayr, L. M., Minor, D. L., Milhollen, M. A., Burgess, M. W. & Kim, P. S. (1996). Identification of D-peptide ligands through mirror-image phage display. *Science,* 271**,** 1854-1857.

Schutte, J., Minna, J. D. & Birrer, M. J. (1989). Deregulated expression of human c-Jun transforms primary rat embryo cells in cooperation with an activated c-Ha-Ras gene and transforms Rat-1A cells as a single gene. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 2257-2261.

Sergeeva, A., Kolonin, M. G., Molldrem, J. J., Pasqualini, R. & Arap, W. (2006). Display technologies: Application for the discovery of drug and gene delivery agents. *Advanced Drug Delivery Reviews,* 58**,** 1622-1654.

Shaulian, E. & Karin, M. (2002). AP-1 as a regulator of cell life and death. *Nature Cell Biology,* 4**,** E131-E136.

Shepherd, N. E., Hoang, H. N., Abbenante, G. & Fairlie, D. P. (2005). Single turn peptide alpha helices with exceptional stability in water. *Journal of the American Chemical Society,* 127**,** 2974-2983.

Shimizu, Y., Kanamori, T. & Ueda, T. (2005). Protein synthesis by pure translation systems. *Methods,* 36**,** 299-304.

Sibrian-Vazquez, M., Jensen, T. J. & Vicente, M. G. H. (2010). Influence of the number and distribution of NLS peptides on the photosensitizing activity of multimeric porphyrin-NLS. *Organic & Biomolecular Chemistry,* 8**,** 1160-1172.

Slutsky, M. M. & Marsh, E. N. G. (2004). Cation-pi interactions studied in a model coiled-coil peptide. *Protein Science,* 13**,** 2244-2251.

Smeal, T., Angel, P., Meek, J. & Karin, M. (1989). Different requirements for formation of Jun-Jun and Jun-Fos complexes. *Genes & Development,* 3**,** 2091-2100.

Smeal, T., Binetruy, B., Mercola, D. A., Birrer, M. & Karin, M. (1991). Oncogenic and transcriptional cooperation with Ha-Ras requires phosphorylation of c-Jun on Serine-63 and Serine-73. *Nature,* 354**,** 494-496.

Smeenk, L. E. J., Dailly, N., Hiemstra, H., van Maarseveen, J. H. & Timmerman, P. (2012). Synthesis of Water-Soluble Scaffolds for Peptide Cyclization, Labeling, and Ligation. *Organic Letters,* 14**,** 1194-1197.

Smith, J. S. & Scholtz, J. M. (1998). Energetics of polar side-chain interactions in helical peptides: Salt effects on ion pairs and hydrogen bonds. *Biochemistry,* 37**,** 33-40.

Smith, L. M., Wise, S. C., Hendricks, D. T., Sabichi, A. L., Bos, T., Reddy, P., Brown, P. H. & Birrer, M. J. (1999). cJun overexpression in MCF-7 breast cancer cells produces a tumorigenic, invasive and hormone resistant phenotype. *Oncogene,* 18**,** 6063-6070.

Steinkruger, J. D., Bartlett, G. J., Woolfson, D. N. & Gellman, S. H. (2012). Strong Contributions from Vertical Triads to Helix-Partner Preferences in Parallel Coiled Coils. *Journal of the American Chemical Society,* 134**,** 15652-15655.

Sun, T.-L., Sun, Y., Lee, C.-C. & Huang, H. W. (2013). Membrane permeability of hydrocarbon-cross-linked peptides. *Biophysical Journal,* 104**,** 1923-32.

Takahashi, R., Hirata, Y., Sakitani, K., Nakata, W., Kinoshita, H., Hayakawa, Y., Nakagawa, H., Sakamoto, K., Hikiba, Y., Ijichi, H., Moses, H. L., Maeda, S. & Koike, K. (2013). Therapeutic effect of c-Jun N-terminal kinase inhibition on pancreatic cancer. *Cancer Science,* 104**,** 337-344.

ThermoScientific. 2016a. *pUC19 DNA* [Online]. Loughborough, UK. Available: https://www.fishersci.com/shop/products/thermo-scientific-puc18-puc19-dna-50-g/fersd0051#, Accessed 15/07/16.

ThermoScientific. 2016b. *Thermo Scientific Phusion Hot Start II High-Fidelity DNA Polymerase* [Online]. Loughborough, UK. Available: https://tools.thermofisher.com/content/sfs/manuals/MAN0012401_Phusion_HotStartII_DNAPolymerase_500U_UG.pdf, Accessed 05/07/16.

ThermoScientific. 2016c. *TOPO® TA Cloning® Kit for Sequencing, with pCR™4-TOPO® Vector* [Online]. Loughborough, UK. Available: https://www.thermofisher.com/order/catalog/product/K457502, Accessed 15/07/16.

Thompson, E. J., Gupta, A., Stratton, M. S. & Bowden, G. T. (2002). Mechanism of action of a dominant negative c-jun mutant in inhibiting activator protein-1 activation. *Molecular Carcinogenesis,* 35**,** 157-162.

Timmerman, P., Beld, J., Puijk, W. C. & Meloen, R. H. (2005). Rapid and quantitative cyclization of multiple peptide loops onto synthetic scaffolds for structural mimicry of protein surfaces. *ChemBioChem,* 6**,** 821-824.

Travisano, P. & Kennan, A. J. (2004). Tuning stability of coiled-coil heterotrimers by selection of steric matching partners. *Organic Letters,* 6**,** 4219-4222.

Trovato, A., Chiti, F., Maritan, A. & Seno, F. (2006). Insight into the structure of amyloid fibrils from the analysis of globular proteins. *PLoS Computational Biology,* 2**,** 1608-1618.

Ullman, C. G., Frigotto, L. & Cooley, R. N. (2011). In vitro methods for peptide display and their applications. *Briefings in Functional Genomics,* 10**,** 125-134.

Van den Brulle, J., Fischer, M., Langmann, T., Horn, G., Waldmann, T., Arnold, S., Fuhrmann, M., Schatz, O., O'Connell, T., O'Connell, D., Auckenthaler, A. & Schwer, H. (2008). A novel solid phase technology for high-throughput gene synthesis. *BioTechniques,* 45**,** 340-343.

Verde, P., Casalino, L., Talotta, F., Yaniv, M. & Weitzman, J. B. (2007). Deciphering AP-1 function in tumorigenesis - Fra-ternizing on target promoters. *Cell Cycle,* 6**,** 2633-2639.

Vincent, T. L., Green, P. J. & Woolfson, D. N. (2013). LOGICOIL-multi-state prediction of coiled-coil oligomeric state. *Bioinformatics,* 29**,** 69-76.

Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R. & Bonovich, M. (2002). Classification of human B-ZIP proteins based on dimerization properties. *Molecular and Cellular Biology,* 22**,** 6321-6335.

Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989). Scissors-Grip Model for DNA recognition by a family of Leucine Zipper Proteins. *Science,* 246**,** 911-916.

Virnekas, B., Ge, L. M., Pluckthun, A., Schneider, K. C., Wellnhofer, G. & Moroney, S. E. (1994). Trinucleotide phosphoramidites - ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Research,* 22**,** 5600-5607.

Vodnik, M., Zager, U., Strukelj, B. & Lunder, M. (2011). Phage Display: Selecting Straws Instead of a Needle from a Haystack. *Molecules,* 16**,** 790-817.

Wagschal, K., Tripet, B. & Hodges, R. S. (1999). De novo design of a model peptide sequence to examine the effects of single amino acid substitutions in the hydrophobic core on both stability and oligomerization state of coiled-coils. *Journal of Molecular Biology,* 285**,** 785-803.

Walensky, L. D., Kung, A. L., Escher, I., Malia, T. J., Barbuto, S., Wright, R. D., Wagner, G., Verdine, G. L. & Korsmeyer, S. J. (2004). Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix. *Science,* 305**,** 1466-1470.

Wick, M., Lucibello, F. C. & Muller, R. (1992). Inhibition of Fos-induced and Ras-induced transformation by mutant Fos proteins with structural alterations in functionally different domains. *Oncogene,* 7**,** 859-867.

Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical Biochemistry,* 179**,** 131-137.

Worrall, J. A. & Mason, J. M. (2011). Thermodynamic analysis of Jun-Fos coiled coil peptide antagonists. *FEBS Journal,* 278**,** 663-72.

Yamada, R. & Kera, Y. (1998). D-amino acid hydrolysing enzymes. *EXS,* 85**,** 145-55.

Yang-Yen, H. F., Chiu, R. & Karin, M. (1990). Elevation of AP-1 activity during F9 cell differentiation is due to increased c-Jun transcription. *New Biologist,* 2**,** 351-362.

Yang, A. S. & Honig, B. (1995). Free-energy determinants of secondary structure formation .1. alpha-helices. *Journal of Molecular Biology,* 252**,** 351-365.

Yap, J. L., Chauhan, J., Jung, K.-Y., Chen, L., Prochownik, E. V. & Fletcher, S. (2012). Small-molecule inhibitors of dimeric transcription factors: Antagonism of protein-protein and protein-DNA interactions. *Medchemcomm,* 3**,** 541-551.

Ye, N., Ding, Y., Wild, C., Shen, Q. & Zhou, J. (2014). Small Molecule Inhibitors Targeting Activator Protein-1 (AP-1). *Journal of Medicinal Chemistry,* 57**,** 6930-6948.

Young, M. R., Li, J. J., Rincon, M., Flavell, R. A., Sathyanarayana, B. K., Hunziker, R. & Colburn, N. (1999). Transgenic mice demonstrate AP-1 (Activator Protein-1) transactivation is required for tumor promotion. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 9827-9832.

Young, T. S. & Schultz, P. G. (2010). Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *Journal of Biological Chemistry,* 285**,** 11039-11044.

Yu, R., Hebbar, V., Kim, D. W., Mandlekar, S., Pezzuto, J. M. & Kong, A. N. T. (2001). Resveratrol inhibits phorbol ester and UV-induced activator protein 1 activation by interfering with mitogen-activated protein kinase pathways. *Molecular Pharmacology,* 60**,** 217-224.

Zeng, X. G., Zhu, H., Lashuel, H. A. & Hu, J. C. (1997). Oligomerization properties of GCN4 leucine zipper ***e*** and ***g*** position mutants. *Protein Science,* 6**,** 2218-2226.

Zhou, N. E., Kay, C. M. & Hodges, R. S. (1992). Synthetic model proteins - positional effects of interchain hydrophobic interactions on stability of 2-stranded alpha-helical coiled-coils. *Journal of Biological Chemistry,* 267**,** 2664-2670.

Zhu, B. Y., Zhou, N. E., Semchuk, P. D., Kay, C. M. & Hodges, R. S. (1992). Design, synthesis and structural characterization of model heterodimeric coiled-coil proteins. *International Journal of Peptide and Protein Research,* 40**,** 171-179.

Zimm, B. H. & Bragg, J. K. (1959). Theory of the phase transition between helix and random coil in polypeptide chains. *Journal of Chemical Physics,* 31**,** 526-535.

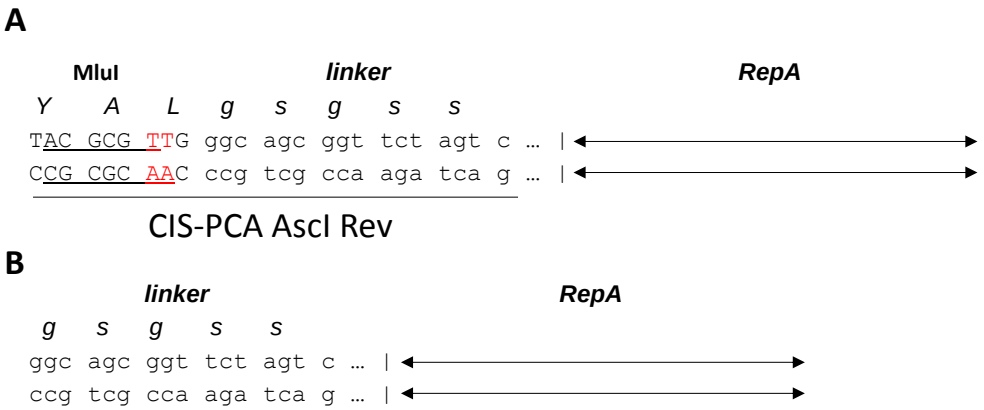
Appendix Contents

Chapter 2 Appendix

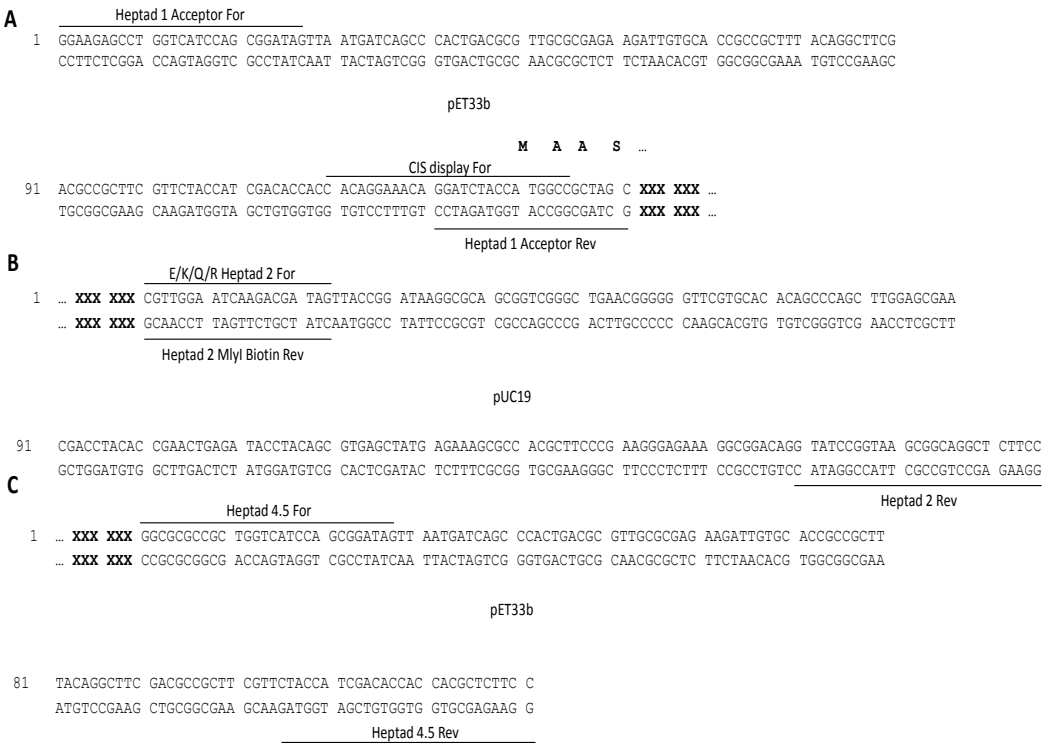
2.1 Test pulldowns of biotin-GG-cJun and SDS-PAGE

To optimise coating quantities of biotin-GG-cJun on Dynal M-280 (Invitrogen, Loughborough, UK) streptavidin-coated magnetic beads, HPLC purified biotin-GG-cJun was diluted for five coating quantities: 0.25 µg, 0.5 µg, 1 µg or 2 µg per 20 µl in PBS. 130 µl of Dynal M-280 beads (25 µl per coating quantity) were washed twice with 500 µl PBS, split between coating quantities, and bead/peptide mixtures were incubated for 15 mins with mixing every 5 mins. After incubation, supernatant was removed and 4X SDS loading buffer (Invitrogen) added to a final concentration of 1X. Beads were washed three times with 500 µl PBS then resuspended in 20 µl PBS/1X SDS loading buffer. Peptide samples were also prepared at 0.25 µg, 0.5 µg, 1 µg or 2 µg per 20 µl PBS/1X SDS loading buffer. Peptide samples, beads after capture, and bead capture supernatants were then run on an SDS-PAGE gel (Appendix Figure 2.4).

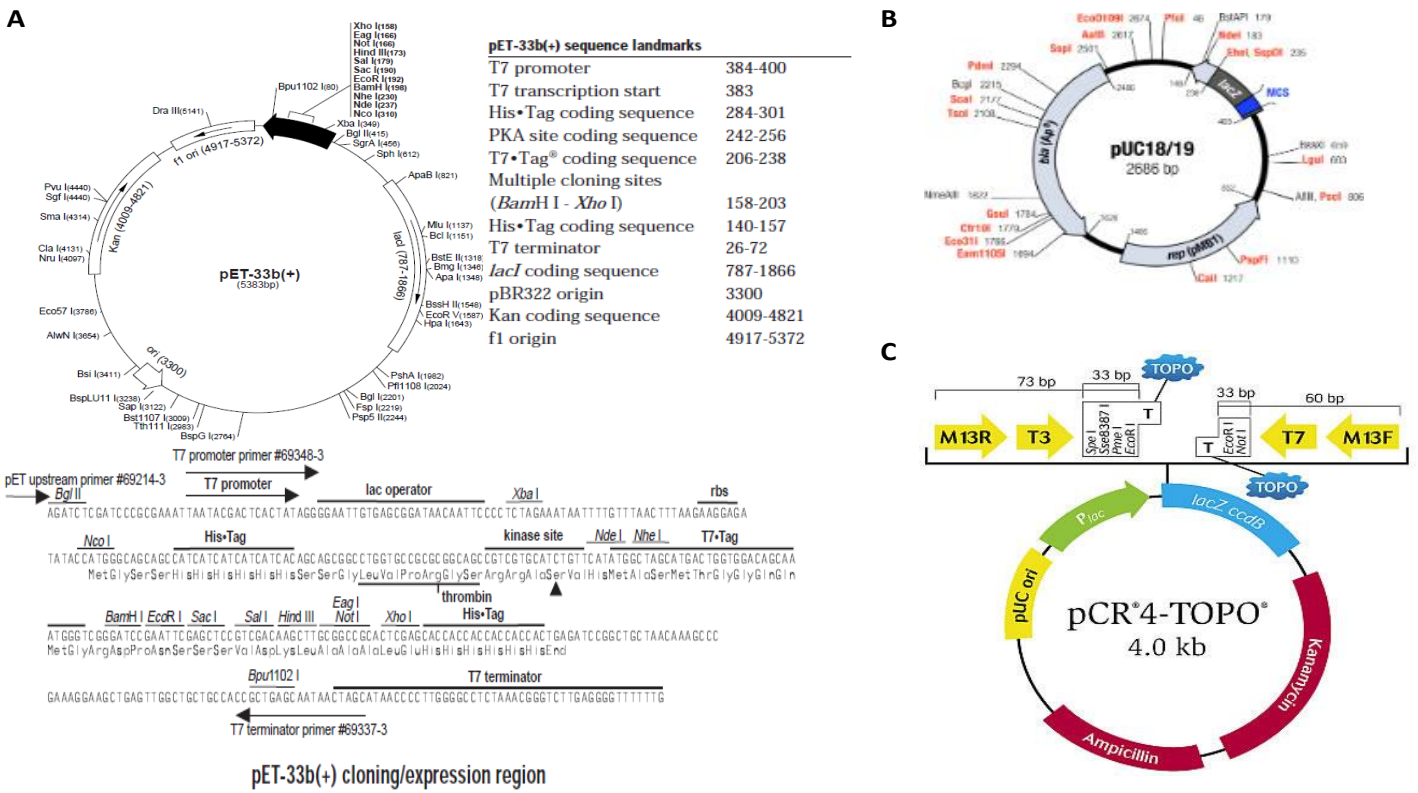
Samples were heated in 20 µl of PBS/1X SDS loading buffer to 95 °C for 10 mins before being loaded onto a 1 mm-thick 4-12 % NuPAGE Novex Bis-Tris polyacrylamide gel (Life Technologies, Loughborough, UK). Gels were run in a Mini-PROTEAN Tetra apparatus (BioRad, Hemel Hempstead, UK) for 35 mins at 180 V with a 1X NuPAGE MES running buffer (50 mM MES, 50 mM Tris-Base, 0.1 % SDS, 1 mM EDTA, pH 7.3; Life Technologies). Gels were stained for ≥one hour in Instant Blue stain solution (Expedeon, Swavesey, UK) before destaining as necessary in ddH_2O.



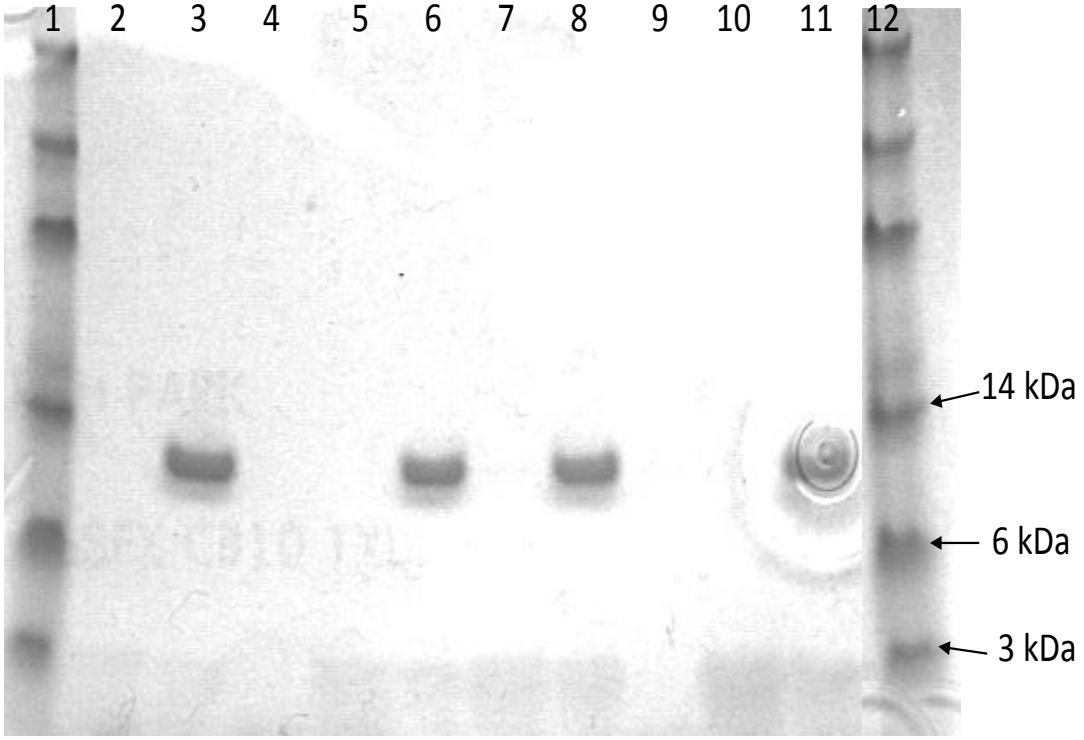
Appendix Figure 2.1: CIS display sequences attached to peptide library variable regions. A (MluI)link-RepA-CIS-Ori sequence attached to ProxiMAX libraries for CIS display screening. The MluI site is underlined, the priming site for reintroduction of an AscI site following CIS display using the CIS-PCA AscI Rev primer to allow PCA screening is marked with a line, and the corresponding two-base mutation is highlighted in red text (TT→CC, Leu →Pro). **B** link-RepA-CIS-Ori sequence attached to 3.4hFosW, 12mer and 16mer libraries for CIS display screening.



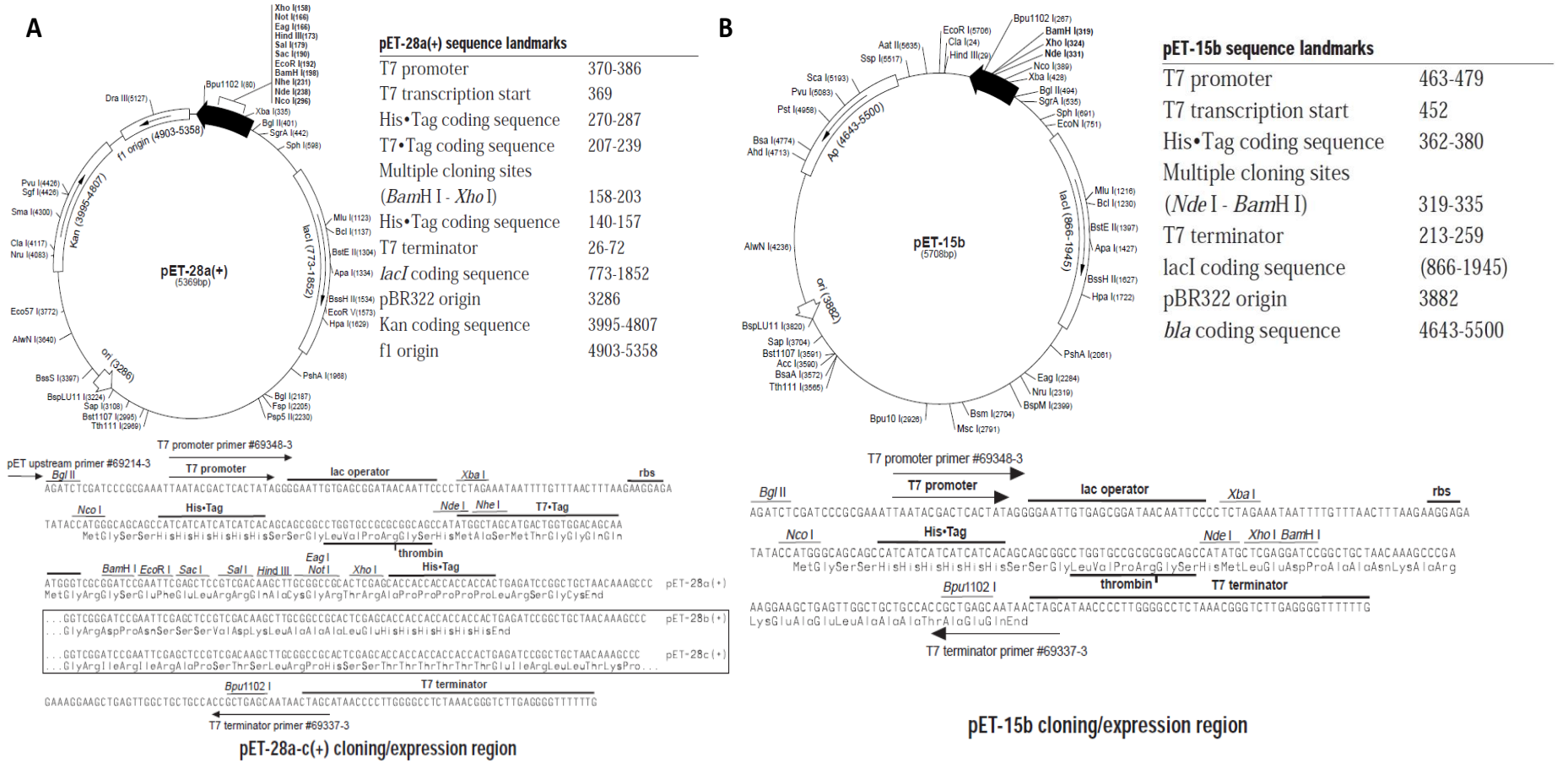
Appendix Figure 2.2: Acceptor sequences used in ProxiMAX library construction. A Heptad 1 Acceptor derived from pET33+ plasmid vector. **B** Heptad 2 Acceptor derived from pUC19 plasmid vector. **C** Heptad 4.5 Acceptor derived from pET33+ plasmid vector. Library codons are blunt-end ligated to acceptors ("XXX") to build Heptads 1, 2 and 4.5. Important priming sites on acceptors are marked with lines.



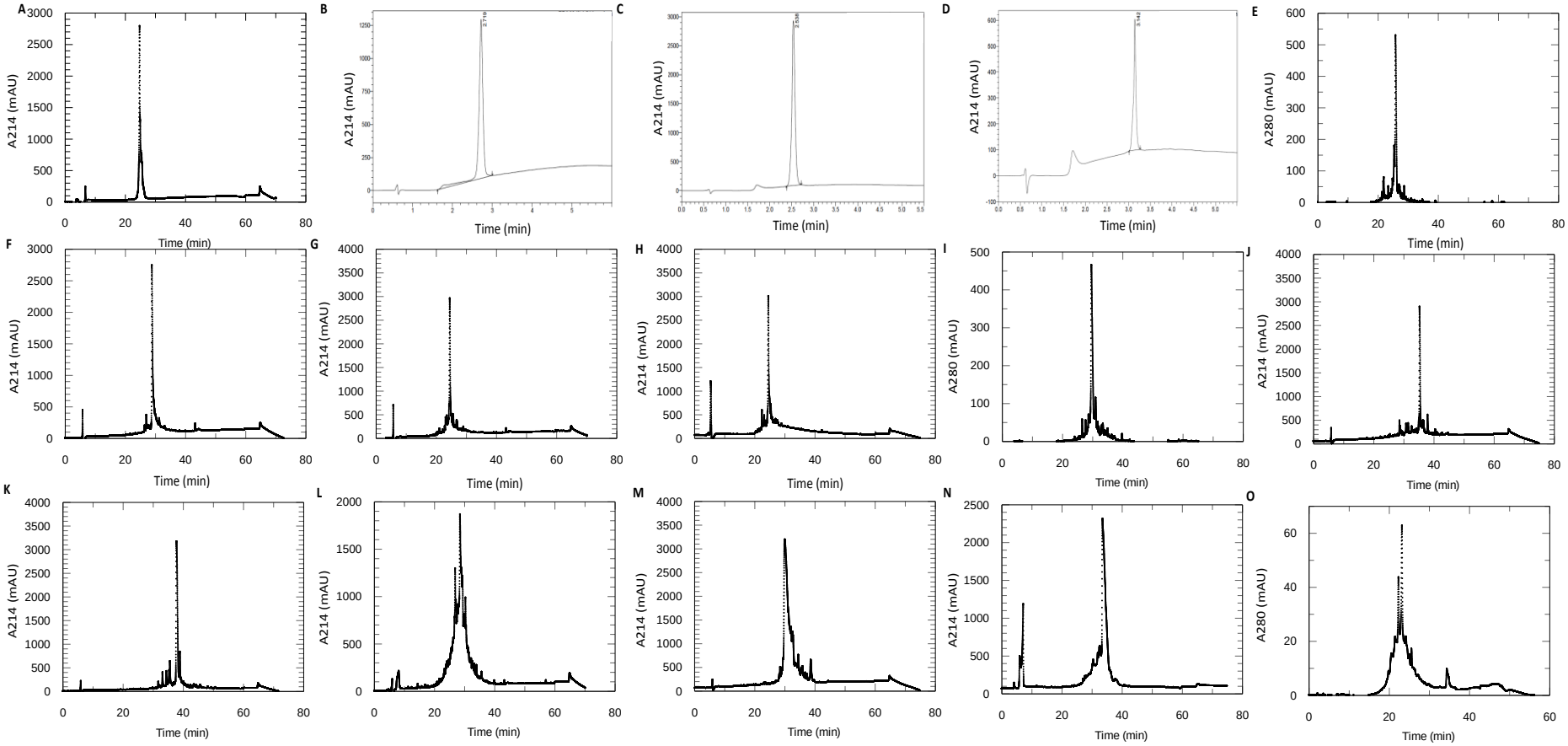
Appendix Figure 2.3: Plasmids from which ProxiMAX Heptad acceptor sequences and RepA for CIS display were amplified. A and **B** pET33+ and pUC19 plasmids from which Heptad 1, 2 and 4.5 Acceptors (see Appendix Figure 2.2) were amplified. **A** pET33+ vector map adapted from source (Novagen, 2016c). **B** pUC19 vector map adapted from source (ThermoScientific, 2016a). **C** pCR4 plasmid from which RepA was amplified to prepare CIS display-ready libraries. Vector map adapted from source (ThermoScientific, 2016c). Plasmid maps show sites for restriction enzyme digestion, relevant Open Reading Frames (such as for antibiotic resistance markers), and in the case of **A** and **C** expanded images of the multiple restriction site regions for sub-cloning and/or protein expression. Primer annealing sites are displayed in **C** as yellow arrows.



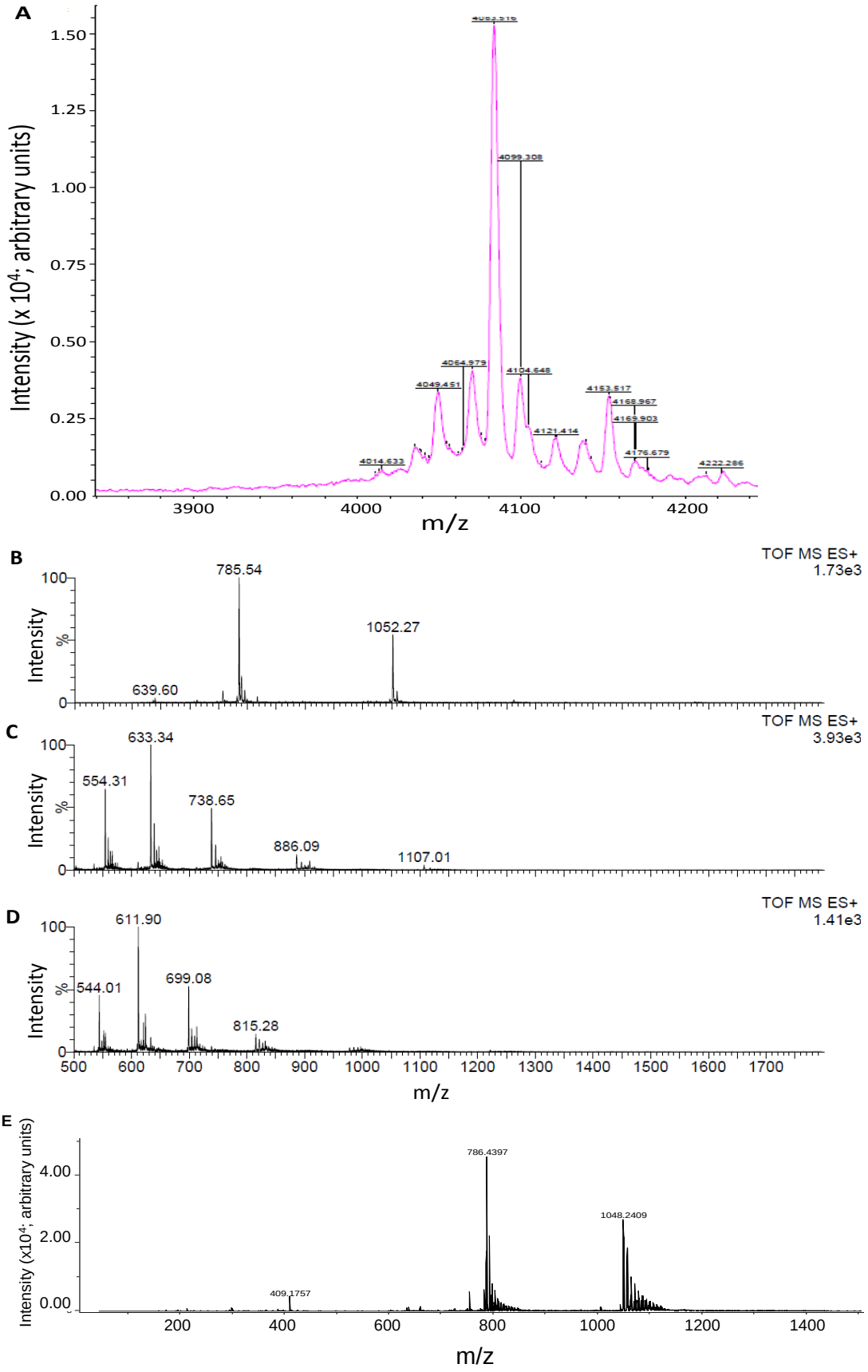
Appendix Figure 2.4: Test pulldown of biotin-GG-cJun peptide on streptavidin-coated magnetic beads for CIS display selections. 4-12% NuPAGE Novex Bis-Tris SDS-PAGE gel (Invitrogen) of biotin-GG-cJun pulled down on streptavidin-coated Dynal M-280 beads (Invitrogen). 20 µl of samples loaded. Bands at ≈12 kDa are denatured streptavidin, bands at ≈3 kDa are biotin-GG-cJun. Lane 1: SeeBlue2 (Invitrogen), Lane 2: 0.25 µg cJun, Lane 3: 0.25 µg cJun bead capture, Lane 4: 0.25 µg cJun capture supernatant, Lane 5: 0.5 µg cJun, Lane 6: 0.5 µg cJun bead capture, Lane 7: 1.0 µg cJun, Lane 8: 1 µg cJun bead capture, Lane 9: 1 µg cJun capture supernatant, Lane 10: 2 µg cJun, Lane 11: 2 µg cJun bead capture, Lane 12: SeeBlue2. Capture of 0.25 µg of cJun on 25 µl beads, and 0.5 µg cJun on 25 µl beads, leads to bands after thermal denaturation and release of captured cJun (Lanes 3 and 6) that match intensities of 0.2 µg and 0.5 µg peptide alone bands (Lanes 2 and 5), indicating that at least 0.5 µg of cJun can be captured on 25 µl beads. Adding 1 µg of cJun to 25 µl beads leads to capture of slightly less than 1 µg based on band intensity after thermal denaturation and release of captured cJun from beads (Lane 8) compared to 1 µg peptide alone (Lane 7), whilst considerably less than 2 µg cJun is captured and released by beads (Lanes 11 and 10). Thus, bead capture capacity was determined to be 0.5 – 1 µg per 25 µl beads, and so CIS display selection stringencies were set at 2 µg per 50 µl beads for R1 “low stringency” selection or 0.5 µg per 25 µl beads for R1 “high stringency” selection.



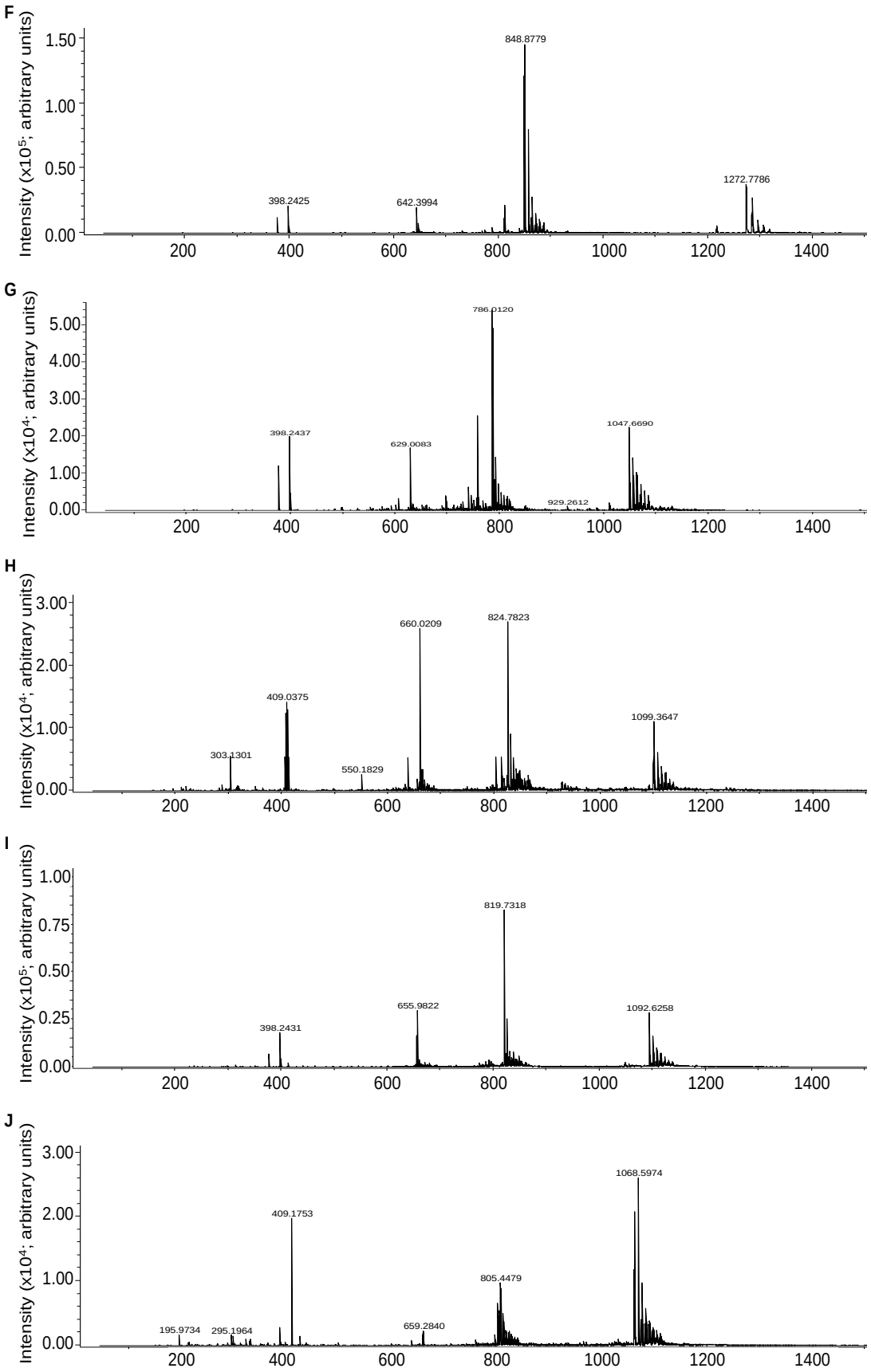
Appendix Figure 2.5: PCA plasmid vector maps. A pET28a+ plasmid map (Novagen). A modified version with additional AscI restriction site, linker and mDHFR2 fragment coding sequences C-terminal of cloned library sequences was used for PCA. Plasmid map adapted from source (Novagen, 2016b). **B** pET15b plasmid map (Novagen). A modified version with additional AscI restriction site, linker and mDHFR1 fragment coding sequences C-terminal of the cloned cJun CC domain was used for PCA. Plasmid map adapted from source (Novagen, 2016a).



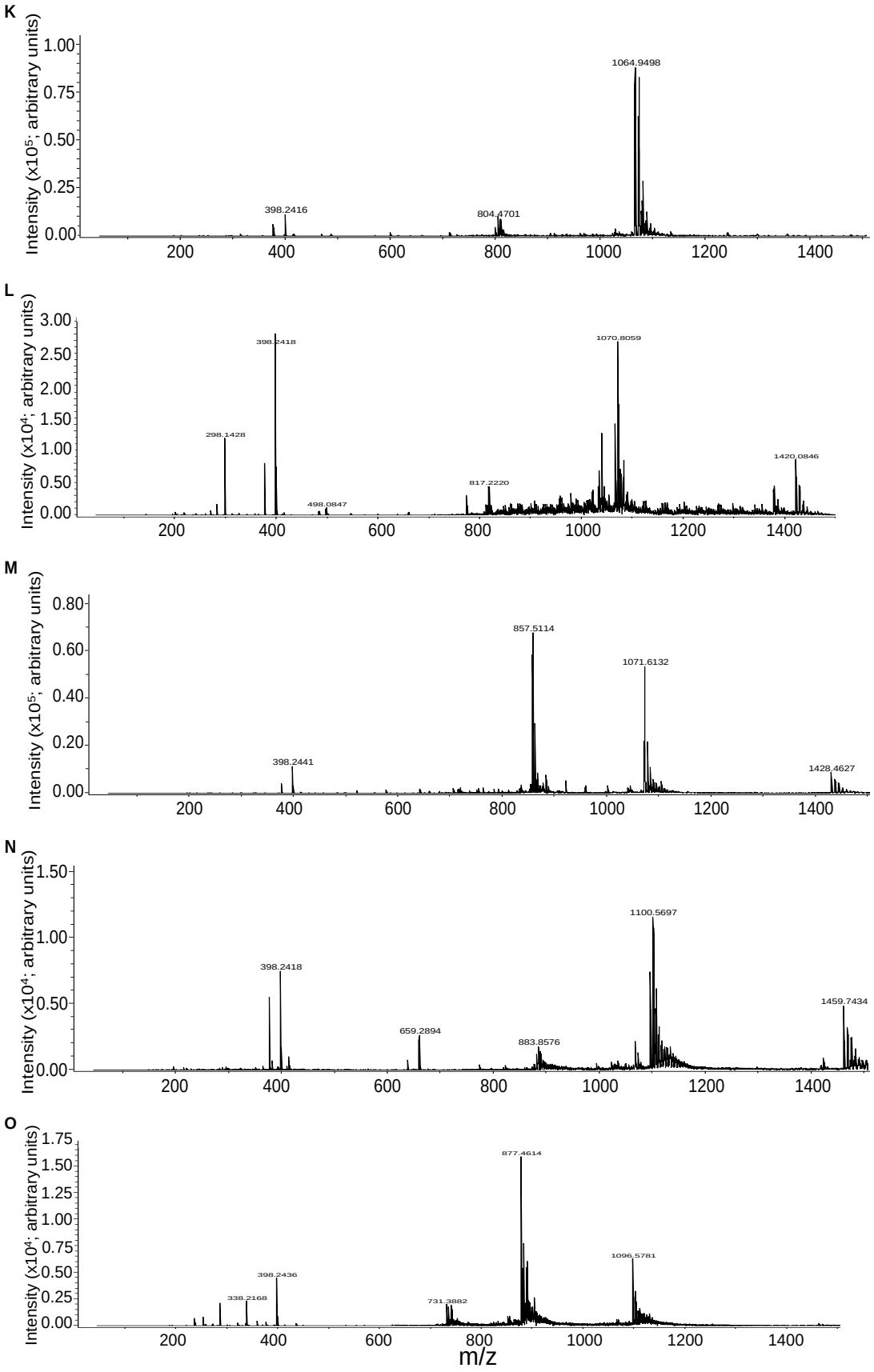
Appendix Figure 2.6: RP-HPLC chromatograms of synthesized peptides. Absorbance at 214 nm or 280 nm (mAU) vs. retention time (min). **A** cJun. **B** Peptide **20**. **C** Peptide **20**-Tat. **D** Peptide **18**HC. **E** LIN20. **F** 2.5pep1. **G** 3.5pep22. **H** 3.5pep23. **I** 3.4HCIS. **J** 3.4LCIS. **K** 3.4W. **L** CIS1. **M** CIS21. **N** CPW. **O** Biotin-GG-cJun.



Appendix Figure 2.7: Mass spectra of HPLC purified synthesized peptides. A MALDI-TOF of cJun. **B**, **C** and **D** ESI-TOF of **20**HC, **20**-Tat and FITC-**20**-Tat. **E** ESI-TOF of LIN20. **Figure continues overleaf.**



Appendix Figure 2.7 continued: Mass spectra of HPLC purified synthesized peptides. F – J ESI-TOF of 2.5pep1, 3.5pep22, 3.5pep23, 3.4HCIS and 3.4LCIS peptides. **Figure continues overleaf.**



Appendix Figure 2.7 continued: Mass spectra of HPLC purified synthesized peptides. K – O ESI-TOF of 3.4W, CIS1, CIS21, CPW and Biotin-GG-cJun peptides.

Chapter 3 Appendix

Appendix Table 3.1: FosW–cJun crystal structure parameters. Parameters are as follows: I/σ(I) denotes signal-to-noise ratio; R_{merge} is the spread of multiple measurements of the same reflections; R_{pim} is the R-factor for redundancy-independent precision of measured diffraction intensities; R_{cryst} , the R-factor, is the difference between the diffraction data and model expressed as a percentage of the diffraction data; R_{free} denotes the R-factor of data not used during refinement; RMSD, the root mean square difference; and B-factors are the temperature factors for mean-square atom displacement from rest position in $Å^2$. *Values given are overall values, values in parentheses relate to the highest resolution measured (1.99 Å). Crystal structure reported to be determined to 2.3 Å corresponding to ≈100 % completion.

Parameters	Values
Indexing and scaling	
Space Group	$I\,4_1\,2_1\,2_1$
a, b, c (Å)	63.82, 63.82, 190.63
α, β, γ (°)	90, 90, 90
N° of merged unique reflections	13418
Resolution (Å) (outer shell)	47.66 - 1.99 (2.04 - 1.99)
Completeness (%)*	96.2 (78.4)
Redundancy*	17.4 (6.3)
I/σ(I)*	19.7 (2.5)
R_{merge} (%)*	9.1 (65.9)
R_{pim} (%)*	2.1 (39.0)
Phasing method	Molecular Replacement
Refinement	
Resolution (Å)	1.99
N° of reflections work (free)	13377 (1180)
R_{cryst}	20.86
R_{free}	24.64
RMSD bond length (Å)	0.007
RMSD bond angle (°)	0.977
N° of atoms	
Protein	1137
Waters	123
Ligands	68
B-factors	
Protein (main chain / side chain)	41.0 / 43.4
Waters	47.2
Ligands	47.6

Appendix Table 3.2: Helix constrained peptide homomeric and heteromeric sample helicities and coiled coil formation. Fractional helicities and 222:208 ratios ($\Theta_{222/208}$, where Θ is ellipticity in mdeg) for peptides alone or an equimolar mixture with cJun as determined by CD spectroscopy at 20 °C. Fractional helicities were calculated according to Equation 4 (see 2.9.6.1). 222:208 ratios are indicative of coiled coil formation where $\Theta_{222/208}$>≈0.9 (Zhou *et al.*, 1992, Lau *et al.*, 1984). All peptides were characterised in KPP buffer (10 mM KH_2PO_4/K_2HPO_4 + 100 mM KF). Values are given to 2 s.f.. Generally, full length constrained peptides that are capable of homodimerisation are also capable of heterodimerisation, whilst the same is not true for constrained, truncated peptides, and peptide homomeric samples are generally more helical than heteromeric samples. Peptides with 222:208 ratios slightly below 0.9 can nevertheless demonstrate dimerisation by CD thermal melt and ITC analyses.

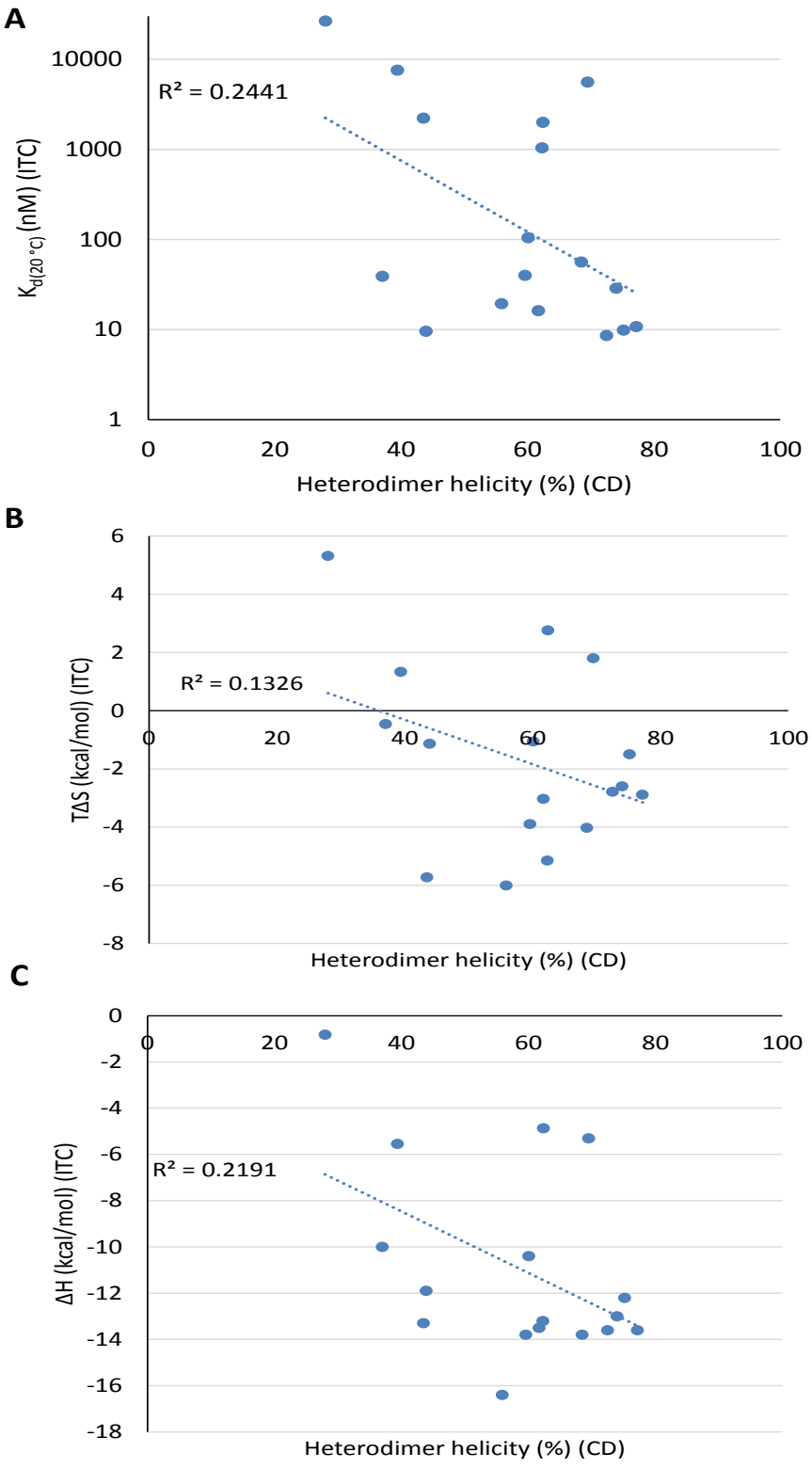
Peptide	Peptide Fractional Helicity (%) (CD)	Peptide $\Theta_{222/208}$ (20°C) (CD)	Peptide–cJun Fractional Helicity (%) (CD)	Peptide–cJun $\Theta_{222/208}$ (20°C) (CD)
FosW	41	1.0	37	1.0
1	44	1.1	44	0.98
2	62	1.1	62	1.0
3	40	0.77	29	0.70
4	59	1.1	60	1.0
5	79	1.1	74	1.1
6	53	1.1	56	1.0
7	85	1.1	73	1.0
8	78	1.1	75	1.0
9	82	1.0	77	1.0
10	64	1.1	34	0.82
11	17	1.1	14	0.61
12	28	0.92	25	0.78
13	38	1.0	27	0.79
14	39	1.0	44	0.97
15	65	1.0	62	1.0
16	48	0.87	49	1.0
17	67	0.94	44	0.86
18	72	1.1	69	1.0
19	64	1.0	60	0.99
20	85	1.1	62	0.99
LIN**20**	50	0.84	39	0.81
LIN**20**-TAT	66	0.96	59	0.90
Pal-**20**	28	1.1	28	0.82
20-TAT	96	1.0	70	0.95
20-NLS-TAT	46	0.81	39	0.79
NLS-TAT	0	-0.25	11	0.31
18HC	40	1.0	44	0.88
20HC	49	1.0	35	0.78

Appendix Table 3.3: Low predicted monomeric helicity does not guarantee poor heterodimer helicity for cJun binders. Sequences of cFos, FosB, Fra1, Fra2, cJun, JunB and JunD from Angel and Karin (1991). cJun, cFos, FosW(E), $FosW_{Core}$, and JunW sequences and heterodimer fractional helicities from Worrall and Mason (2011). $JunW_{CANDI}$ sequence and heterodimer fractional helicity from Rao *et al.* (2013). 4hFosW sequence and heterodimer fractional helicity from Crooks *et al.* (2011). $FosW_{CANDI}$ sequence and heterodimer fractional helicity from Mason *et al.* (2007b). *AS and GAP helix capping motifs added for Agadir analysis. Heterodimer helicities for FosB, Fra1, Fra2, JunB and JunD have not been determined (ND), however Agadir monomeric helicities allow comparison with other peptides. Values are given to 2 s.f.. †Peptide–cFos fractional helicities: cFos-binding JunW and $JunW_{CANDI}$ sequences, also displaying low monomeric helicities relative to heterodimer fractional helicities, have been included for comparison with cJun binders.

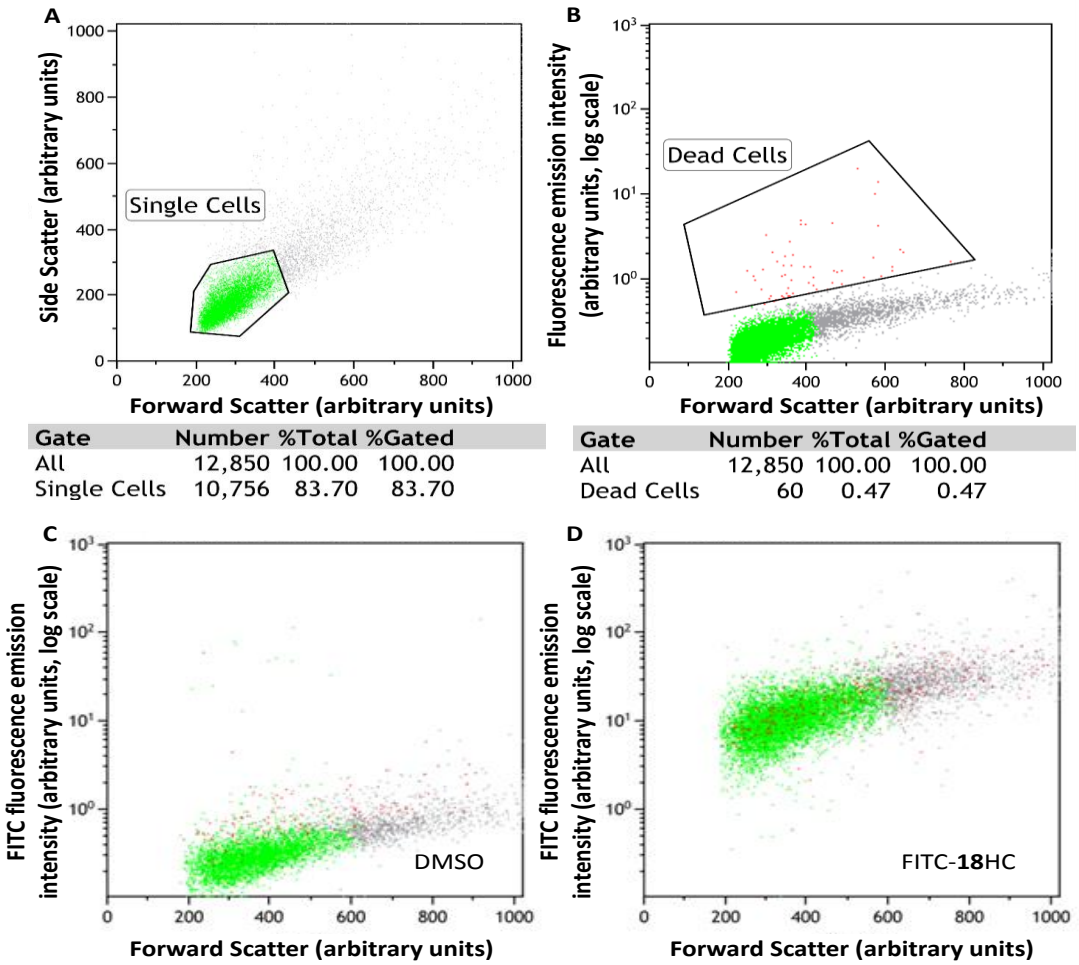
Peptide Name	Peptide Sequence	Monomeric Helicity (%) (Agadir)*	Peptide–cJun Fractional Helicity (%) (CD)
FosW	LDELQAEIEQLEERNYALRKEIEDLQKQLEKL	26	37
4hFosW	LDELQREIEQLEELNYALQKEIEDLQKQ	22	60
$FosW_{CANDI}$	LDELQAEIEQLEDQNYALQKEVEDLRKELEKL	22	54
JunW	AAELEERVKTLKAEIYELQSEANMLREQIAQL	17	46†
$JunW_{CANDI}$	AAELEERAKTLKAEIYELRSKANMLREHIAQL	22	42†
$FosW_{Core}$	IDELQAEVEQLEERNYALRKEVEDLQKQAEKL	10	43
FosW(E)	LDELEAEIEQLEEENYALEKEIEDLEKELEKL	7.9	88
cFos	TDTLQAETDQLEDEKYALQTEIANLLKEKEKL	3.5	28
FosB	TDRLQAETDQLEEEKYELESEIAELQKEKERL	5.0	ND
Fra1	TDFLQAETDKLEDEKYGLQREIEELQKQKERL	14	ND
Fra2	TEKLQAETEELEEEKYGLQKEIAELQKEKEKL	4.1	ND
cJun	IARLEEKVKTLKAQNYELASTANMLREQVAQL	3.7	14
JunB	IARLEDKVKTLKAENYGLSSTAGLLREQVAQL	2.4	ND
JunD	ISRLEEKVKTLKSQNYELASTASLLREQVAQL	2.6	ND

Appendix Table 3.4: Low predicted monomeric helicity does not guarantee poor heterodimer helicity for constrained cJun binders. Fractional helicities for heterodimers of peptides with cJun by CD (Equation 4, see 2.9.6.1) are compared to expected monomeric helicities. These were estimated by dividing the number of residues that are between those forming the constraint, including the residues forming the constraint themselves, by the total number of residues in the peptide, to represent the helicity of the peptide arising from locking the residues between those of the constraint in an α-helical conformation. This approach was adopted as the Agadir algorithm can only predict monomeric helicities of linear peptides (Lacroix *et al.*, 1998). This approach estimates the minimal helicity of monomeric peptides, not considering propagation of helicity along the peptide from constrained regions (Harrison *et al.*, 2010). Differences between measured heteromeric helicity and estimated monomeric helicities are also given. Values are given to 2 s.f..

Peptide	Peptide–cJun Fractional Helicity (%) (CD)	Estimated monomeric helicity (%)	Difference in helicity between measured and predicted
1	44	16	28
2	62	16	46
3	29	16	13
4	60	16	44
5	74	16	58
6	56	31	25
7	73	31	41
8	75	31	44
9	77	31	46
10	34	46	-12
11	14	68	-54
12	25	46	-21
13	27	68	-42
14	44	40	3.5
15	62	60	2.3
16	49	40	8.9
17	44	46	-1.8
18	69	35	34
19	60	35	26
20	62	40	22



Appendix Figure 3.1: Poor correlation between peptide K_d values, entropy or enthalpy of cJun binding and heterodimer helicity. Values calculated for cFos, FosW and peptides **1**-**20** (excluding **3**), **20**-TAT and **20**-NLS-TAT only. **A** K_d values (log scale) vs. heterodimer helicity. **B** Entropy of binding (TΔS) vs. heterodimer helicity. **C** Binding enthalpy (ΔH) vs. heterodimer helicity.



Gate	Number	%Total	%Gated
All	12,850	100.00	100.00
Single Cells	10,756	83.70	83.70

Gate	Number	%Total	%Gated
All	12,850	100.00	100.00
Dead Cells	60	0.47	0.47

Appendix Figure 3.2: Representative example of uptake of internalising FITC-peptide by MCF-7 cell population. FACS dot plots displaying MCF-7 cell population picking (gating) of single live cells (**A** and **B**), and uptake of a representative example peptide FITC-**18**HC (**D**) vs. DMSO negative control (**C**) in gated cells. **A** Side Scatter (SS) as a measure of cell granularity vs. Forward Scatter (FS) corresponding to the size of cells. Single MCF-7 cells (green dots, within black lines) were separated from other species (grey dots). Gated cell percentages relative to total cell population are given below figure panel. **B** 7-aminoactinomycin D fluorescence emission intensity vs. FS. From the single cells gated in **A**, those displaying 7-aminoactinomycin D fluorescence (cell membrane disruption) and thus representing dead cells (red dots, within black lines) were separated from live cells (low 7-aminoactinomycin D fluorescence), which were gated for analysis of peptide uptake vs. DMSO control in **C** and **D**. Gated cell percentages relative to total cell population are given below figure panel. In **C** and **D**, FITC fluorescence emission intensity is plotted vs. FS. Grey dots represent species other than single live cells that were not counted to determine cell population uptake of FITC-peptide. **C** DMSO negative control. The FITC intensity of single live cells (green dots) is negligible. **D** Cells treated with FITC-**18**HC. Approx. 98 % of cells contained FITC-**18**HC, with 2 % either dead (red dots), not single cells (grey dots), or having FITC fluorescence similar to that of the DMSO control and so not internalising peptide.

Chapter 4 Appendix

Appendix Table 4.1: Top 20 peptide sequences from deep sequencing of the 3.5h library CIS display output. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 3.5h library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides Values are given to 2 s.f. except observation frequencies (raw number). **Table continues overleaf.**

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
Low stringency selection, round R3	1	IRKLELEIEAIEAELLEIRAQLEEIQLQVIQL	32	19237	19
	2	VQEIELQLEELEKQNKLIQEQLIAVQKEIEQV	32	8437	8.2
	3	IAKLEKRIIELQIKLLQAQVEKVQLEIKAVEEQVRLV	37	8128	7.9
	4	IQEIELEILALEAQLAEIQIQLIRIQKQVLKLQVQQL	37	5163	5.0
	5	IKLVQREIAAWRIKNQQLRIEIRQLEAI	28	4809	4.7
	6	VAAIELEVQELELQNRIIRAQLAIVQAEVEKL	32	4583	4.4
	7	LIQLEAQLEEIEIELKQIKLRLEEVRRKLRQLEKQRQV	38	4303	4.2
	8	IEQLEAEVKKLEIQLEEIRLQIIELEKQIKQI	32	3823	3.7
	9	VLAIEKELEEIELRNAALQLQIQEWQEQIAKLQKLQRL	38	3255	3.2
	10	IEELEIQIQELEEQLRQIQKQLKAIEAQVLAY	32	2759	2.7
	11	VQKLELEIAKLERELKRIRQQLQEVEKQLKKL	32	2540	2.5
	12	VQQLEIEIEEVEKEIQLVRQEIAQLQAQVEAL	32	1982	1.9
	13	IEEIQQQIQALQKRNKELRIEILKLRILRLEKELARY	37	1862	1.8
	14	VAKLEIELEIIEIQLQLIREQLKELQKQIQAV	32	1797	1.7
	15	VEALEAEIQRLEKELQEIREQLEELREKVLEL	32	1781	1.7
	16	IELQLEIEALRLLLIKKQIQELQKKVAQI	29	1700	1.6
	17	IQKLEKEVLILEQQLIEVRAQIEQIQQQVAILQAQLAIV	39	1620	1.6
	18	VQKLELEIEQLEQELLLIQAQLKKVELELKRL	32	1478	1.4
	19	VQIIKAEVRKVKKENRQLRIQIAELRLQKRLAEF	34	1311	1.3
	20	VKLLEAELQAIEAQLQQIREEIEELQKQVQELKQERWKKI	40	1286	1.2
Low stringency selection, round R4	1	VQEIELQLEELEKQNKLIQEQLIAVQKEIEQV	32	22242	22
	2	IKLVQREIAAWRIKNQQLRIEIRQLEAI	28	13207	13
	3	IRKLELEIEAIEAELLEIRAQLEEIQLQVIQL	32	13148	13
	4	VAAIELEVQELELQNRIIRAQLAIVQAEVEKL	32	8624	8.4
	5	VLAIEKELEEIELRNAALQLQIQEWQEQIAKLQKLQRL	38	6747	6.6
	6	IAKLEKRIIELQIKLLQAQVEKVQLEIKAVEEQVRLV	37	3939	3.8
	7	IQEIELEILALEAQLAEIQIQLIRIQKQVLKLQVQQL	37	2722	2.6
	8	IEQLEAEVKKLEIQLEEIRLQIIELEKQIKQI	32	1920	1.9
	9	VQQLEIEIEEVEKEIQLVRQEIAQLQAQVEAL	32	1905	1.9
	10	VQIIKAEVRKVKKENRQLRIQIAELRLQKRLAEF	34	1731	1.7
	11	LIQLEAQLEEIEIELKQIKLRLEEVRRKLRQLEKQRQV	38	1648	1.6
	12	IELQLEIEALRLLLIKKQIQELQKKVAQI	29	1472	1.4
	13	IEELEIQIQELEEQLRQIQKQLKAIEAQVLAY	32	1375	1.3
	14	IEEIQQQIQALQKRNKELRIEILKLRILRLEKELARY	37	1336	1.3
	15	VQAIEQELEALEIQNERLRQELLIIQIQVQAL	32	1262	1.2
	16	VQKLELEIAKLERELKRIRQQLQEVEKQLKKL	32	1242	1.2
	17	VEQLEKEVEELQQEIELIERQFKEIEKQILLL	32	1189	1.2
	18	VEALEAEIQRLEKELQEIREQLEELREKVLEL	32	964	0.9
	19	VAKLEIELEIIEIQLQLIREQLKELQKQIQAV	32	957	0.9
	20	IALELQLEEIRLAIIQAQLEQLQKEVEKL	29	946	0.9

Appendix Table 4.1 continued: Top 20 peptide sequences from deep sequencing of the 3.5h library CIS display output. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 3.5h library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides Values are given to 2 s.f. except observation frequencies (raw number).

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
High stringency selection, round R3	1	VRIVELQIVRIQNARIQLELKRVRQRILAL	30	1014	7.0
	2	IEIIKQQIIIVQARIRAIQLQVKLVRKQVLQV	32	913	6.3
	3	IKKIELQLLEKRIQIVRLKIQI	22	781	5.4
	4	VQIVKIEIKIIEKRIKIIQAQIQAI	25	536	3.7
	5	IAIVRLENIELRLKLRRLQIQIRLVQKQILQV	32	432	3.0
	6	FKKFELQVLLIKKKNLILRIQDQRIEVKIVRRLRIQVARI	40	431	3.0
	7	IKRIEIQLEPQRIQVKIVELELRAI	25	336	2.3
	8	YAIIKLQVRLVQRRNQIVRLKVKKV	25	317	2.2
	9	IEEVQIQIKRVLRLELQVRAIKIRLVEKQNQEVRLQLAKV	40	309	2.1
	10	VQLVQKRIIIIRLRNIKVELQLLEV	25	301	2.1
	11	VIKLQLRLQVRKIRIQVQLVRKKILKI	27	296	2.1
	12	IIIVKRRIIQLQAQIKALQLEVKLVRKQILIV	32	252	1.7
	13	VLKLKARIQIVRRRSHVLIQLEREIIRL	28	239	1.7
	14	VEIVKLEIEIVRKQVRIIQAEILAL	25	234	1.6
	15	VRILQILKIQNALVRAEIKLVREQVIKV	28	232	1.6
	16	IIKIEARIIALRLRNLRVRIVQKEIKQIKVEIL	33	221	1.5
	17	VEIVKEQIIIIQIENKLLQIRLVRKRVIKIEQELAIL	37	220	1.5
	18	IQIVKRRIAIIQLQNIQLRLKQLQIEVIQI	30	212	1.5
	19	IIEVRQKLLIVQRIQLQNAQWRLEVKIIKKQVLRL	35	200	1.4
	20	YRIVEIQIRLVQKKVRIVQLRLEQV	25	195	1.4
High stringency selection, round R4	1	VRIVELQIVRIQNARIQLELKRVRQRILAL	30	8389	16
	2	IEIIKQQIIIVQARIRAIQLQVKLVRKQVLQV	32	7460	14
	3	IKKIELQLLEKRIQIVRLKIQI	22	6110	11
	4	VQIVKIEIKIIEKRIKIIQAQIQAI	25	1622	3.0
	5	IKRIEIQLEPQRIQVKIVELELRAI	25	1329	2.5
	6	IAIVRLENIELRLKLRRLQIQIRLVQKQILQV	32	1320	2.4
	7	VQLVQKRIIIIRLRNIKVELQLLEV	25	1044	1.9
	8	VIKLQLRLQVRKIRIQVQLVRKKILKI	27	1040	1.9
	9	IIEVRQKLLIVQRIQLQNAQWRLEVKIIKKQVLRL	35	1025	1.9
	10	IIKIEARIIALRLRNLRVRIVQKEIKQIKVEIL	33	962	1.8
	11	FKKFELQVLLIKKKNLILRIQDQRIEVKIVRRLRIQVARI	40	917	1.7
	12	IIIVKRRIIQLQAQIKALQLEVKLVRKQILIV	32	811	1.5
	13	VRLLQRRIIEIQVRELRIQFRLLKEEVLAL	30	758	1.4
	14	IEEVQIQIKRVLRLELQVRAIKIRLVEKQNQEVRLQLAKV	40	755	1.4
	15	VEIVKEQIIIIQIENKLLQIRLVRKRVIKIEQELAIL	37	709	1.3
	16	YAIIKLQVRLVQRRNQIVRLKVKKV	25	669	1.2
	17	VLKLKARIQIVRRRSHVLIQLEREIIRL	28	636	1.2
	18	VAKYKIKLRIQAQVREVRLQVQLVQQWEKQVAEV	34	636	1.2
	19	YRIVEIQIRLVQKKVRIVQLRLEQV	25	591	1.1
	20	VLLIKLLRKEIIIIQAQVKQLRLELKLVEKQLLII	35	544	1.0

Appendix Table 4.2: Overall composition of peptide sequences from deep sequencing of the 3.5h library CIS display outputs. The percentages of total sequence reads of 3.5h peptides identified as having been selected by low or high stringency CIS display after selection rounds R3 or R4 that were observed ≥50 times, 2 – 10 times, or only once. The numbers of sequences observed the stated number of times are also given in parentheses, and the total number of sequence reads is given in the rightmost column. Percentages given to 2 s.f.. The percentages of sequences observed ≥50 times is higher for the low stringency selection than the high stringency condition, which may indicate more successful enrichment of particular peptides in this selection (noting the fewer sequence reads for the former). This is supported by the substantial proportion of the library observed 2 – 10 times relative to the peptides observed only once in this selection stringency. The similar percentages of peptides observed 2 – 10 times for the high stringency condition relative to the low stringency condition corroborates the suggestion that the high stringency condition was not enriching peptides efficiently, as a lesser proportion of peptides were enriched above ten copies than the low stringency condition.

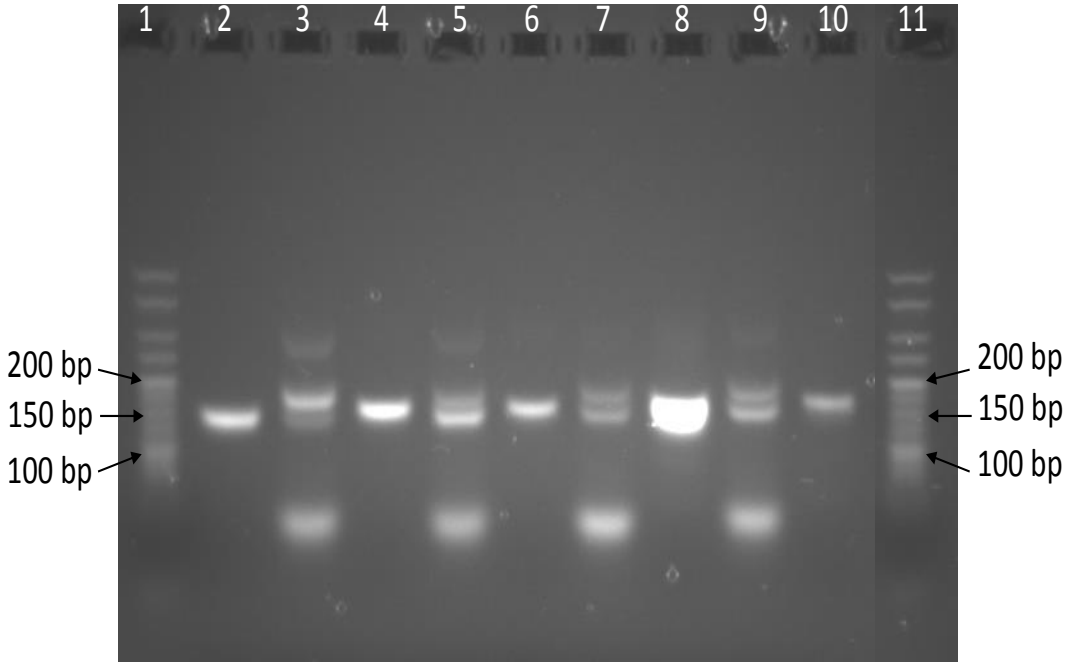
Selection stringency and round	Sequences observed ≥50 times: percentage (number)	Sequences observed 2 – 10 times: percentage (number)	Sequences observed once: percentage (number)	Total number of sequence reads
Low stringency, round R3	0.26 (110)	37 (16,022)	62 (26,629)	43,087
Low stringency, round R4	0.33 (85)	23 (6048)	76 (19,606)	25,884
High stringency, round R3	0.09 (98)	28 (30,262)	71 (77,675)	108,952
High stringency, round R4	0.18 (131)	24 (16,952)	75 (54,045)	71,814

Appendix Table 4.3: Homomeric and heteromeric sample helicities and coiled coil formation for peptides selected from the 3.5h library. Fractional helicities and 222:208 ratios ($\Theta_{222}/_{208}$, where Θ is ellipticity in mdeg) for peptides alone or an equimolar mixture with cJun as determined by CD spectroscopy at 20 °C. Fractional helicities were calculated according to Equation 4 (see 2.9.6.1). 222:208 ratios are indicative of coiled coil formation where $\Theta_{222}/_{208}$>≈0.9 (Zhou *et al.*, 1992, Lau *et al.*, 1984). FosW determined in KPP buffer (10 mM KH_2PO_4/K_2HPO_4 + 100 mM KF), CIS1, CIS21 and CPW determined in "low salt" buffer (10 mM KH_2PO_4/K_2HPO_4). Values are given to 2 s.f.. FosW, CIS21 and CPW form homodimeric coiled coils (CCs) as well as heterodimeric CCs, whilst CIS1 CC formation via self-interaction or interaction with cJun is debatable. Helicity of heteromeric samples is generally slightly lower than for homomeric samples.

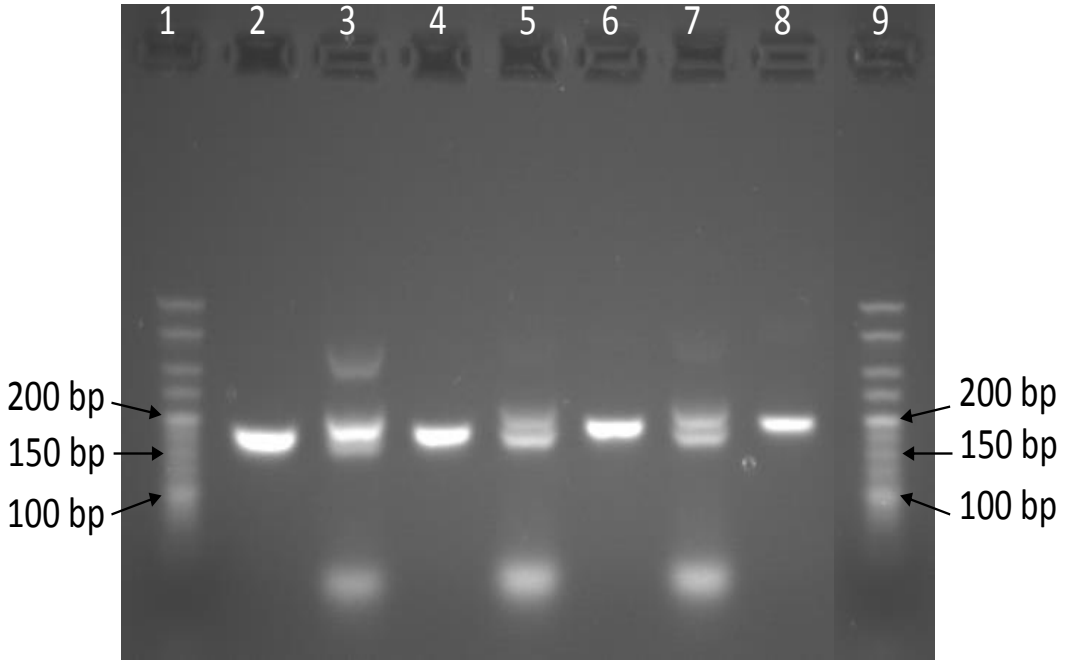
Peptide	**Peptide Fractional Helicity (%) (CD)**	**Peptide $\Theta_{222}/_{208}$ (20°C) (CD)**	**Peptide–cJun Fractional Helicity (%) (CD)**	**Peptide–cJun $\Theta_{222}/_{208}$ (20°C) (CD)**
FosW	41	1.0	37	1.0
CIS1	31	0.69	31	0.71
CIS21	56	0.97	48	0.92
CPW	62	1.0	53	0.95

Appendix Table 4.4: CPW is predicted to have preference for trimerisation rather than dimerisation. *In silico* analysis of oligomeric state propensity using the Logicoil algorithm (Vincent *et al.*, 2013). Logicoil outputs scores for probability of adoption of oligomeric states (column headings); scores are ranked with the most positive score as the most likely, and the least positive score the least likely. Logicoil can only predict oligomeric states of homomeric species, with preference for heteromeric complexes to adopt certain oligomeric state extrapolated from homomeric preferences. Values are given to 2 s.f..

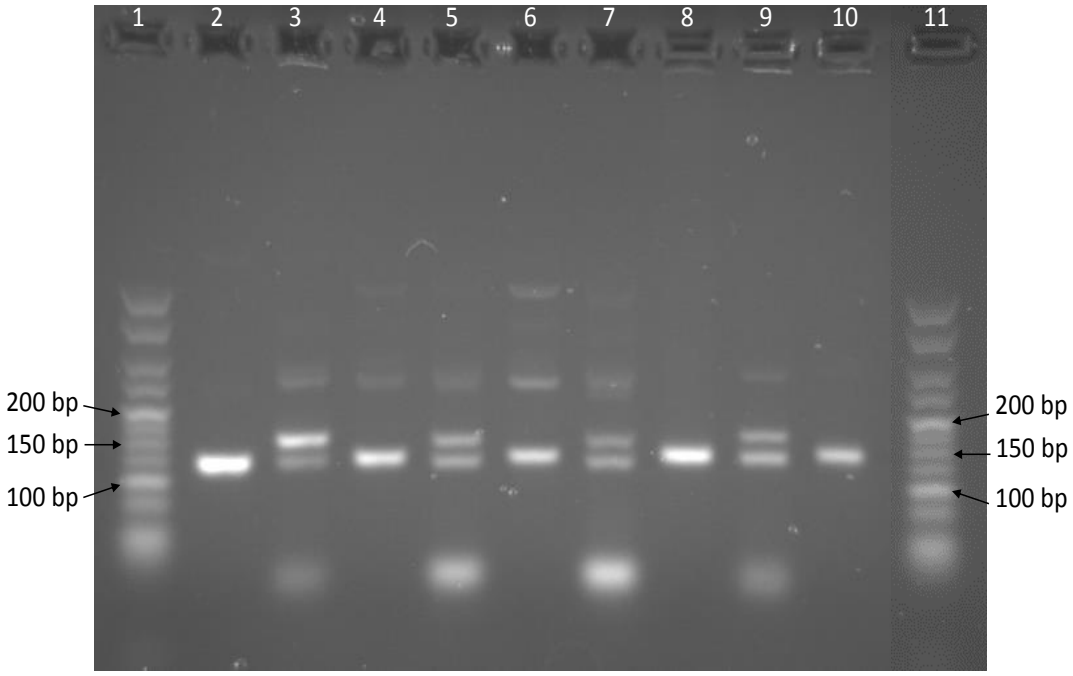
Selection	Peptide	Sequence	Parallel Dimer	Antiparallel Dimer	Trimer	Tetramer
3.5h CIS display low stringency	CPW	VQEIEQEIQELEKRIKQIQQEFQEIEQQIALL	0.91	1.0	1.3	0.92



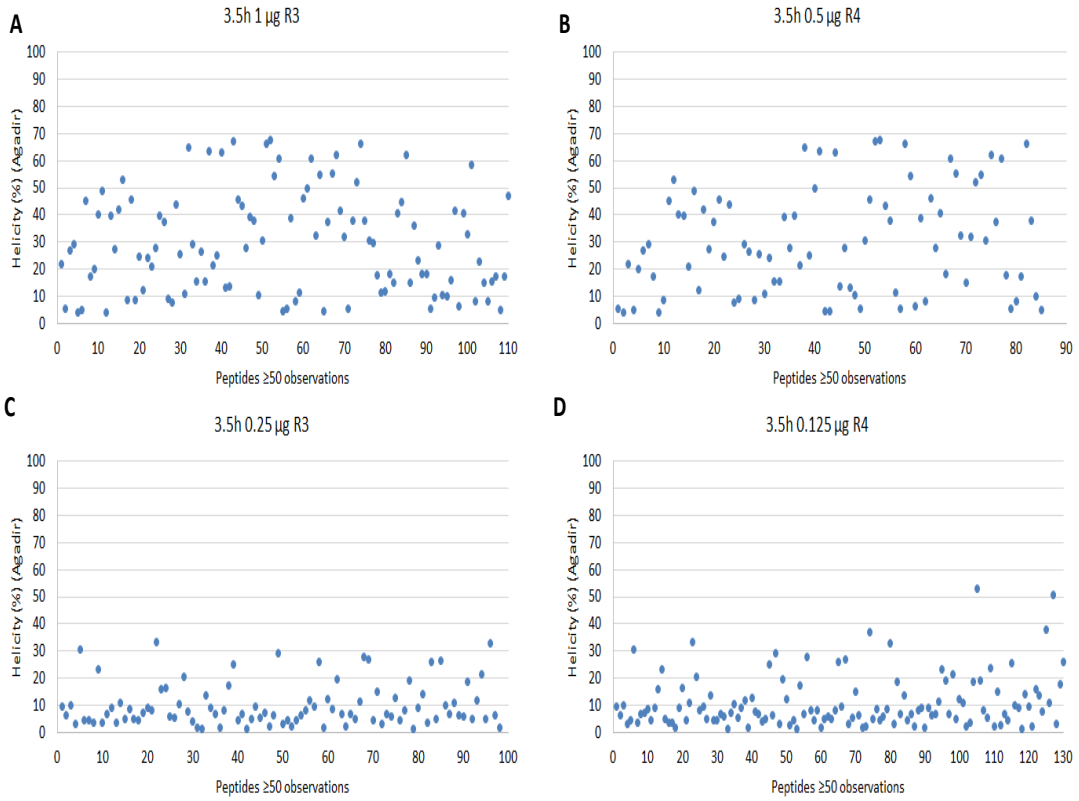
Appendix Figure 4.1: Heptad 1 cassette build. Agarose gel of Heptad 1 cassette build. Samples and ladders run on 2% agarose/Tris-Borate EDTA gel stained with SybrGold. Expected sizes of desired species are given in parentheses. Lane 1: Hyperladder V (Bioline), Lane 2: Heptad 1 Acceptor (151 bp), Lane 3: Heptad 1 C1 ligation, Lane 4: Heptad 1 C1 MlyI digest (157 bp), Lane 5: Heptad 1 C2 ligation, Lane 6: Heptad 1 C2 MlyI digest (160 bp), Lane 7: Heptad 1 C3 ligation, Lane 8: Heptad 1 C3 MlyI digest (166 bp), Lane 9: Heptad 1 C4 ligation, Lane 10: Heptad 1 C4 MlyI digest (173 bp), Lane 11: Hyperladder V. Ligation lanes consist of unligated acceptor band (lowest band, same size as acceptor pre-ligation), successful ligation product (higher MW band), and formation of some undesirable concatamers (highest MW band).



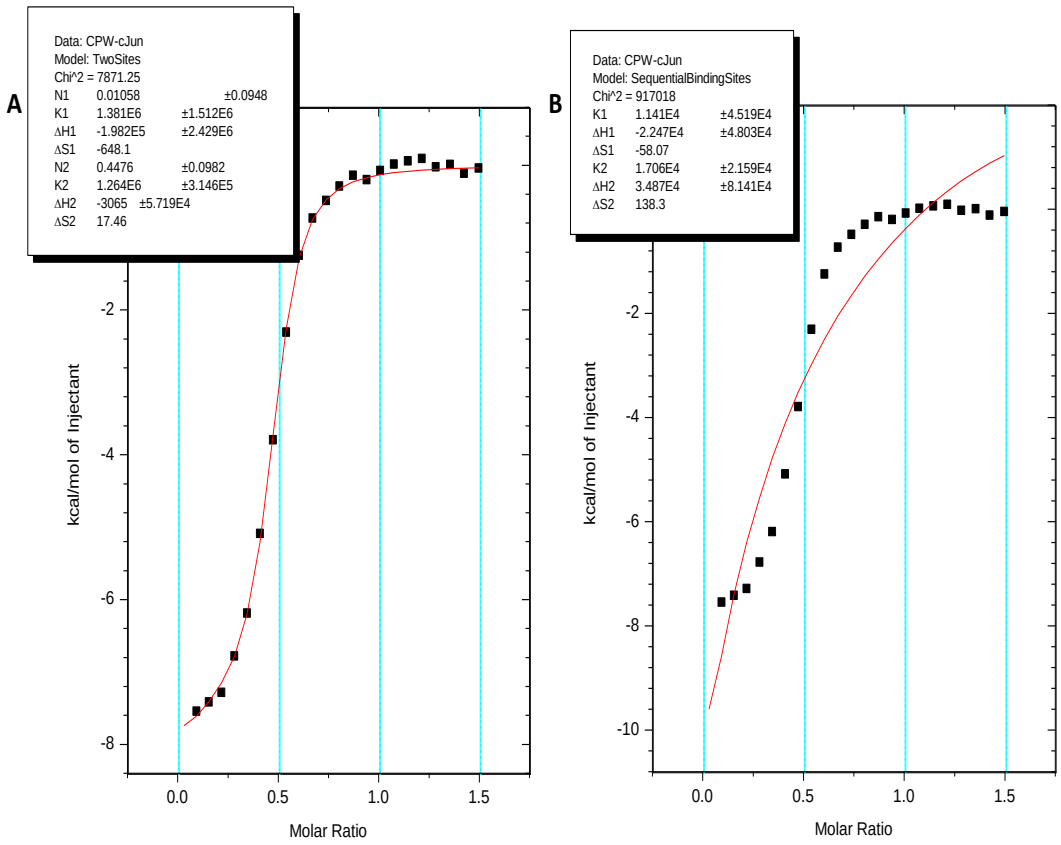
Appendix Figure 4.2: Heptad 2 cassette build. Agarose gel of Heptad 2 cassette build. Samples and ladders run on 2% agarose/Tris-Borate EDTA gel stained with SybrGold. Expected sizes of desired species are given in parentheses. Lane 1: Hyperladder V (Bioline), Lane 2: Heptad 2 Acceptor (194 bp), Lane 3: Heptad 2 C1 ligation, Lane 4: Heptad 2 C1 MlyI digest (200 bp), Lane 5: Heptad 2 C2 ligation, Lane 6: Heptad 2 C2 MlyI digest (206 bp), Lane 7: Heptad 2 C3 ligation, Lane 8: Heptad 2 C3 MlyI digest (212 bp), Lane 9: Hyperladder V. Ligation lanes consist of unligated acceptor band (lowest band, same size as acceptor pre-ligation), successful ligation product (higher MW band), and formation of some undesirable concatamers (highest MW band).



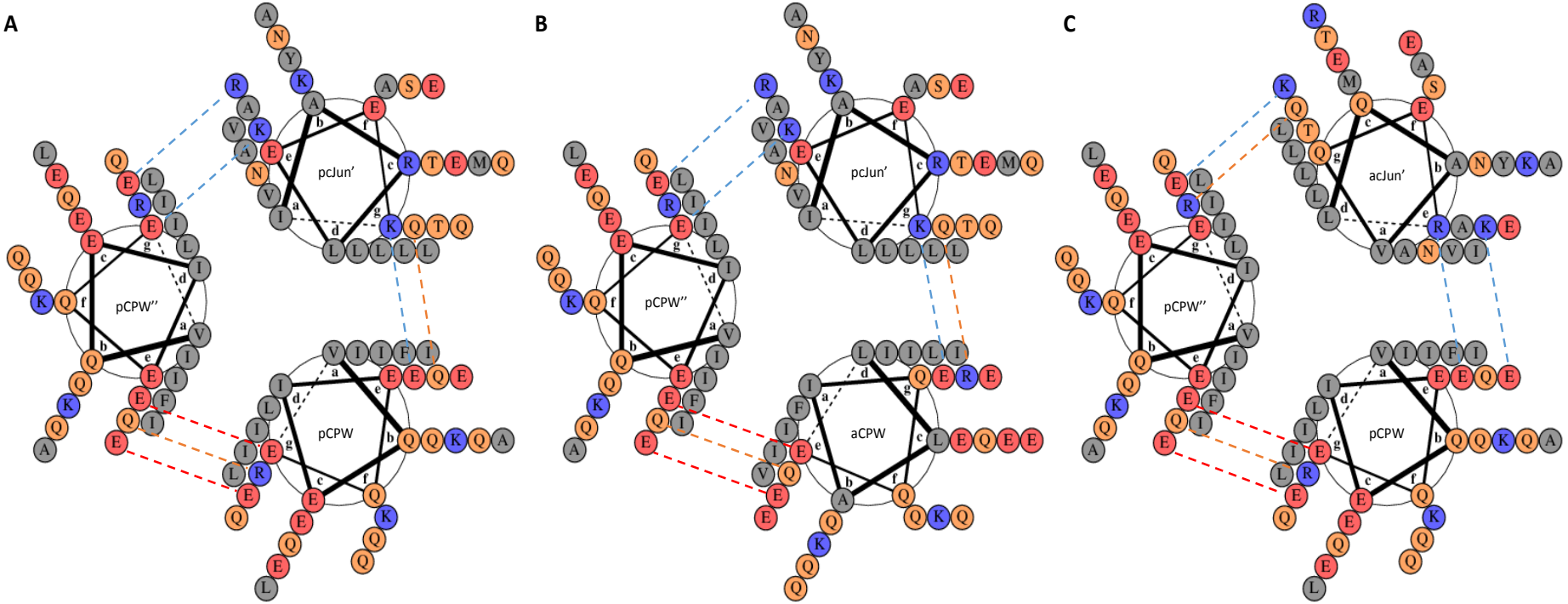
Appendix Figure 4.3: Heptad 4.5 cassette build. Agarose gel of stages of Heptad 4.5 cassette build. Samples and ladders run on 2% agarose/Tris-Borate EDTA gel stained with SybrGold. Expected sizes of desired species are given in parentheses. Lane 1: Hyperladder V (Bioline), Lane 2: Heptad 4.5 Acceptor (131 bp), Lane 3: Heptad 4.5 C1 ligation, Lane 4: Heptad 4.5 C1 MlyI digest (134 bp), Lane 5: Heptad 4.5 C2 ligation, Lane 6: Heptad 4.5 C2 MlyI digest (137 bp), Lane 7: Heptad 4.5 C3 ligation, Lane 8: Heptad 4.5 C3 MlyI digest (140 bp), Lane 9: Heptad 4.5 C4 ligation, Lane 10: Heptad 4.5 C4 MlyI digest (143 bp), Lane 11: Hyperladder V (Bioline). Ligation lanes consist of unligated acceptor band (lowest band, same size as acceptor pre-ligation), successful ligation product (higher MW band), and formation of some undesirable concatamers (final highest MW band).



Appendix Figure 4.4: CPW is less helical than the average peptide selected by CIS display low stringency selection. Helicity of peptides observed ≥50 times in the peptide subset selected by CIS display low stringency (**A** and **B**) and high stringency (**C** and **D**) R3 and R4 selections as calculated by Agadir (Lacroix *et al.*, 1998). CPW has an Agadir helicity of 15 %, making it half as helical as other peptides enriched by CIS display in the low stringency selection, and more similar to the average helicity of peptides enriched by the less successful high stringency selection.



Appendix Figure 4.5: ITC data for CPW–cJun does not fit well to two site binding models for a 3-state transition. A Fitting of "CIS→PCAWinner" (CPW) isothermal titration calorimetry (ITC) data for binding to cJun to a binding model featuring two distinct, independently-bound binding sites. **B** Fitting of CPW–cJun ITC data to a binding model featuring binding of one molecule to one site to facilitate binding of the second molecule at the second binding site. The unacceptably large errors in all parameters for both hypothetical separate or sequential binding events indicates the inapplicability of these models to the CPW–cJun data, and the applicability instead of the one site binding model for a 2-state transition presented in Figure 4.6.



Appendix Figure 4.6: Possible heterotrimeric CPW–cJun complexes (legend continues on next page). Helical wheel diagrams visualised using DrawCoil (Grigoryan, 2016) of three possible heterotrimer arrangements of CPW–CPW''–cJun', that will require future verification. A heterotrimer featuring two CPW helices and one cJun helix is suggested from similar SEC peak intensities arising from cJun absorbance in the CPW:cJun 1:1 stoichiometric mix sample at retention times for CPW–cJun oligomer and leftover cJun (see Figure 4.8), and the arguably greater propensity for CPW residues, rather than cJun residues, to adopt trimers. The most favourable side chain rotamer for Ile favours acute packing such as that found in trimers (Harbury *et al.*, 1993, Harbury *et al.*, 1994). Selection of Ile at six out of ten ***a*** and ***d*** positions in CPW would therefore strongly encourage formation of a trimer over a dimer. At the remaining three ***a*** and ***d*** positions, two Leu residues (***d***2 and ***d***4.5) would encourage a dimer (Harbury *et al.*, 1993), whilst ***a***1 Val would encourage trimerisation (Harbury *et al.*, 1994). **Legend continues overleaf.**

Appendix Figure 4.6 continued: Possible heterotrimeric CPW–cJun complexes. In the possible heterotrimers above, at least 17 – 18 out of 27 residues favour acute packing (63 – 66%), with aromatic Phe also likely to encourage higher order oligomers than dimers to enable its inclusion in the core away from solvent (Lovejoy *et al.*, 1993, Travisano and Kennan, 2004). Interhelical electrostatic interactions between helices also show some stabilisation of heterotrimers, though repulsions that could disfavour such complexes are also evident. Naming conventions are: first helix–second helix''–third helix', where the first helix is depicted in the bottom right of this Figure, the second helix is depicted on the left, and the third helix is depicted top right. Core interhelical ***a–a'*** and ***d–d'*** interactions are not shown for clarity. Stabilising (blue for expected, orange for potentially occurring) and repulsive (red) interhelical $\boldsymbol{e}_{i}$–$\boldsymbol{g'}_{i-7}$/$\boldsymbol{g}_{i}$–$\boldsymbol{e'}_{i+7}$ and $\boldsymbol{e}_{i}$–$\boldsymbol{e'}_{i-7}$/$\boldsymbol{g}_{i}$–$\boldsymbol{g'}_{i+7}$ interactions are shown as straight dashed lines. Possible intrahelical interactions between outerface ***b***, ***c*** and ***f*** residues stabilising the CPW or cJun helices are also omitted for clarity; see Figure 4.7 for drawing of these on a CPW helix. Residues are coloured grey for hydrophobic, red for positively charged, orange for polar, and blue for negatively charged. **A** The first scenario, pCPW–pCPW''–pcJun', would generate an ***a***4 '***a*** layer' consisting of F–F–A, similar to W–W–L that has been observed in the up-up-down homotrimer of the Coil-Ser peptide (Lovejoy *et al.*, 1993). **B** The second scenario, aCPW–pCPW''–pcJun', would generate at ***a***2 '***a*** layer' of F–I–V and a '***d*** layer' of I–F–L, similar to an '***a*** layer' of Phe–Ala–Ala in a model peptide (Travisano and Kennan, 2004). **C** The third scenario, pCPW–pCPW''–acJun', would generate an F–F–L ***a***4 '***a*** layer' analogous to that in the first scenario. All three scenarios feature three possible stabilising attractive electrostatic interhelical interactions with two possible repulsive interactions.

Chapter 5 Appendix

Appendix Table 5.1: Top 20 peptide sequences from deep sequencing of the 12mer library CIS display outputs. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 12mer library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides. Values are given to 2 s.f. except observation frequencies (raw number). **Table continues overleaf.**

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
Low stringency selection, round R3	1	SYDVTITITLKM	12	196	11
	2	VTIFVNITNTWQ	12	169	9.7
	3	TNIIINVIVQFQ	12	150	8.6
	4	IVTYFKFTIHYV	12	91	5.2
	5	AITIIIITTHDP	12	88	5.1
	6	THWMYIHIQIVQ	12	80	4.6
	7	FFIDIQVTFHYQ	12	78	4.5
	8	MATINISWEIVW	12	73	4.2
	9	DTYITIEVQVYV	12	71	4.1
	10	VKIYVNIHIQAG	12	71	4.1
	11	YTVMHIFYFSYL	12	67	3.8
	12	YIHITYYTIVIE	12	66	3.8
	12	YVMFEMHVYYSK	12	62	3.6
	14	QKTFIFEIHIVL	12	59	3.4
	15	MFTIYIYIGWQG	12	56	3.2
	16	TSYSYIFVNIQM	12	55	3.2
	17	IVYVFIDFKTWV	12	53	3.0
	18	IEFYINIVYVET	12	53	3.0
	19	KIEIHIYMISPK	12	51	2.9
	20	YHNVYINVHIYQ	12	51	2.9
Low stringency selection, round R4	1	SYDVTITITLKM	12	715	5.7
	2	TNIIINVIVQFQ	12	641	5.2
	3	VTIFVNITNTWQ	12	464	3.7
	4	AITIIIITTHDP	12	391	3.1
	5	VKIYVNIHIQAG	12	236	1.9
	6	IVTYFKFTIHYV	12	233	1.9
	7	YVMFEMHVYYSK	12	197	1.6
	8	DTYITIEVQVYV	12	178	1.4
	9	THWMYIHIQIVQ	12	168	1.3
	10	FFIDIQVTFHYQ	12	164	1.3
	11	KIEIHIYMISPK	12	154	1.2
	12	YSIHAEYIIIAV	12	145	1.2
	13	VYIIVEFFGYSQ	12	144	1.2
	14	QKTFIFEIHIVL	12	139	1.1
	15	YIHITYYTIVIE	12	139	1.1
	16	YTVMHIFYFSYL	12	139	1.1
	17	EYFTIIIITHKQ	12	134	1.1
	18	MATINISWEIVW	12	132	1.1
	19	IYYKYISIHVTV	12	128	1.0
	20	MFTIYIYIGWQG	12	122	1.0

Appendix Table 5.1 continued: Top 20 peptide sequences from deep sequencing of the 12mer library CIS display outputs. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 12mer library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides. Values are given to 2 s.f. except observation frequencies (raw number).

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
High stringency selection, round R3	1	KMSIFIQINFSK	12	76	12
	2	YHIKIEVNFHFP	12	72	12
	3	QVNIYFTTTIEK	12	67	11
	4	IYIFININGWNN	12	65	11
	5	SIDISIYFYGWS	12	63	10
	6	YFYTSIYIDIYWY	14	59	9.5
	7	IHIFIISHNTWP	12	59	9.5
	8	EFIEFKFIFYVQ	12	55	8.9
	9	MHSFYIIINIEV	12	52	8.4
	10	TNIIIHIEFIME	12	50	8.1
High stringency selection, round R4	1	YHIKIEVNFHFP	12	648	1.5
	2	KMSIFIQINFSK	12	565	1.3
	3	MHSFYIIINIEV	12	516	1.2
	4	IYIFININGWNN	12	515	1.2
	5	YFYTSIYIDIYWY	13	449	1.0
	6	SFDTIEIIFIVV	12	418	1.0
	7	TNIIIHIEFIME	12	402	0.9
	8	QFKIEFYIYYYK	12	387	0.9
	9	QVNIYFTTTIEK	12	384	0.9
	10	TISITFVYIASG	12	379	0.9
	11	SIDISIYFYGWS	12	376	0.9
	12	IYFEIYIYMSTG	12	374	0.9
	13	EFIEFKFIFYVQ	12	357	0.8
	14	KMHIYHHIYVHI	12	355	0.8
	15	YIHIEFKISYSQ	12	324	0.7
	16	YHIKIEIWTTSW	12	323	0.7
	17	LNYYFFNFQIVI	12	305	0.7
	18	YNIIISLEMQFI	12	303	0.7
	19	TIIFINVFHSWH	12	298	0.7
	20	VKLYITINFQHE	12	297	0.7

Appendix Table 5.2: Top 20 peptide sequences from deep sequencing of the 16mer library CIS display outputs. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 16mer library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides. Values are given to 2 s.f. except observation frequencies (raw number). **Table continues overleaf.**

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
Low stringency selection, round R3	1	KKINMFYTEVYIVIIE	16	503	6.9
	2	TYIIFIDITQHSQYYV	16	203	2.8
	3	SYTFEMFVVVYEHVNS	16	184	2.5
	4	YHSTIIIIHPYVTYVI	16	164	2.3
	5	GNIIIFVFIKNDFHQH	16	154	2.1
	6	YHHWVHIELHIIHNKP	16	152	2.1
	7	IYVYYTYSIIKYHHIW	16	152	2.1
	8	KISYIITWWYVEFKPK	16	132	1.8
	9	VHLIEQHATFINIYIM	16	129	1.8
	10	TEHYHIKITVVFANRQ	16	120	1.7
	11	IFVYFSSYNYKPQYFR	16	119	1.6
	12	VHHNFIVNFVFIIQDQ	16	115	1.6
	13	YFIFYTTEIMIVTMGQ	16	111	1.5
	14	MHSYYEVFSFSIQVTA	16	105	1.5
	15	KNFIKETWKYIIIIVQ	16	104	1.4
	16	HKYSKTEKTHIIFIWY	16	102	1.4
	17	YHIKVKFHVHWTVEFP	16	100	1.4
	18	IHHTFMFYHFIHIQVD	16	99	1.4
	19	SYIKYSFEMRIEWVF	15	97	1.3
	20	TITIVYYEIAKRQFPP	16	95	1.3
Low stringency selection, round R4	1	KKINMFYTEVYIVIIE	16	1772	8.3
	2	TYIIFIDITQHSQYYV	16	546	2.5
	3	SYTFEMFVVVYEHVNS	16	459	2.1
	4	YHHWVHIELHIIHNKP	16	345	1.6
	5	YHSTIIIIHPYVTYVI	16	315	1.5
	6	VHHNFIVNFVFIIQDQ	16	301	1.4
	7	GNIIIFVFIKNDFHQH	16	270	1.3
	8	VHLIEQHATFINIYIM	16	257	1.2
	9	KISYIITWWYVEFKPK	16	256	1.2
	10	IYVYYTYSIIKYHHIW	16	253	1.2
	11	TEHYHIKITVVFANRQ	16	227	1.1
	12	YFIFYTTEIMIVTMGQ	16	223	1.0
	13	SFEYTITIKVIYPIDI	16	220	1.0
	14	MHSYYEVFSFSIQVTA	16	219	1.0
	15	IHHTFMFYHFIHIQVD	16	203	0.9
	16	YHIKVKFHVHWTVEFP	16	192	0.9
	17	YDHTYTYFFIAFTVVK	16	188	0.9
	18	KNFIKETWKYIIIIVQ	16	188	0.9
	19	TITIVYYEIAKRQFPP	16	181	0.8
	20	THVHTIIITYHSWEKH	16	181	0.8

Appendix Table 5.2 continued: Top 20 peptide sequences from deep sequencing of the 16mer library CIS display outputs. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 16mer library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides. Values are given to 2 s.f. except observation frequencies (raw number).

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
High stringency selection, round R3	1	KFTFTIIMQFNNHYFL	16	266	6.1
	2	VTITIMFWESQVQIAV	16	188	4.3
	3	VHHTMSYYYITIVYFP	16	178	4.1
	4	VTHTKFIFFHIIHMPA	16	167	3.8
	5	KVQHYSFTMFSIIAWR	16	144	3.3
	6	TRTTHTIFITISYSIA	16	122	2.8
	7	VHFTFINIYHLEEQGF	16	113	2.6
	8	AINIEISITFFFRYKP	16	106	2.4
	9	YFQIEHHYFIFHVQNQ	16	104	2.4
	10	NFSSMYSIFIVINVGS	16	104	2.4
	11	QARYKITFFSFSFHSA	16	102	2.3
	12	TFYSYIFIIQPDWPSR	16	102	2.3
	13	HYKNTQHIISIVIISG	16	97	2.2
	14	TYKFSIIIVHAAEQTW	16	94	2.2
	15	KIDSYFIFIYTATIIP	16	93	2.1
	16	HYFEYTHTFSIFYIEW	16	81	1.9
	17	RYNYFIEITWVFMQPQ	16	79	1.8
	18	KTYYHITTHVTYNITL	16	78	1.8
	19	KVEYIIIIQQRTFHQT	16	78	1.8
	20	LKTHTIEFTISINVND	16	77	1.8
High stringency selection, round R4	1	KFTFTIIMQFNNHYFL	16	589	5.3
	2	VHHTMSYYYITIVYFP	16	507	4.5
	3	VTITIMFWESQVQIAV	16	434	3.9
	4	KVQHYSFTMFSIIAWR	16	321	2.9
	5	VHFTFINIYHLEEQGF	16	270	2.4
	6	VTHTKFIFFHIIHMPA	16	267	2.4
	7	INIEISITFFFRYKP	15	256	2.3
	8	NFSSMYSIFIVINVGS	16	232	2.1
	9	KTFSTIHVEFTVMFVD	16	221	2.0
	10	YFQIEHHYFIFHVQNQ	16	210	1.9
	11	QARYKITFFSFSFHSA	16	194	1.7
	12	TFYSYIFIIQPDWPSR	16	177	1.6
	13	KIDSYFIFIYTATIIP	16	173	1.5
	14	HYKNTQHIISIVIISG	16	170	1.5
	15	YTTHKITFQVVNIQII	16	143	1.3
	16	RYNYFIEITWVFMQPQ	16	136	1.2
	17	KTRSYTWSQHVIVIFI	16	133	1.2
	18	LKTHTIEFTISINVND	16	132	1.2
	19	TYKFSIIIVHAAEQTW	16	131	1.2
	20	TRTTHTIFITISYSIA	16	127	1.1

Appendix Table 5.3: Top 20 peptide sequences from deep sequencing of the 2.5h library CIS display outputs. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 2.5h library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides. Values are given to 2 s.f. except observation frequencies (raw number). **Table continues overleaf.**

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
Low stringency selection, round R3	1	VKAVREQLIIIQIKLLEL	18	607	3.0
	2	VEKVQIRIQLLQLQVIAL	18	500	2.5
	3	LIKLELQVQIREVRLQLKAL	20	436	2.1
	4	VLKIEIVLEI	10	433	2.1
	5	VLAIEIRLLRQRVIQV	16	402	2.0
	6	IKLLQQRNIIIQIEQLEIKVLEV	23	381	1.9
	7	IIQLKLQLKEIEIQLKIY	18	336	1.6
	8	IQELEIRLAIIRQEIQII	18	245	1.2
	9	IIQVRLKVIQLRRQLKRV	18	205	1.0
	10	VEKIEIQVRRVRIKILEV	18	195	1.0
	11	VAQLEIKIIKLQVQRL	16	189	0.9
	12	IIKIKIELRLIRQQHQIL	18	189	0.9
	13	IAIIELELQQVRLKVKAI	18	174	0.9
	14	IKLRIKQEVERLELQIQIV	19	167	0.8
	15	IIRIEIQLQAKIARH	15	163	0.8
	16	VLKIQLQLKIIQLQVEQL	18	158	0.8
	17	VLQVRIEILEVRKEVKQI	18	154	0.8
	18	IEEIQLRLKLRHKRIQIQLIQV	22	153	0.7
	19	VARIEIKIELIEQKLQEI	18	151	0.7
	20	IIKIKLRLAEWQLKIIRV	18	150	0.7
Low stringency selection, round R4	1	VKAVREQLIIIQIKLLEL	18	1645	2.3
	2	LIKLELQVQIREVRLQLKAL	20	1500	2.1
	3	IKLLQQRNIIIQIEQLEIKVLEV	23	1473	2.0
	4	VEKVQIRIQLLQLQVIAL	18	1350	1.8
	5	IIQLKLQLKEIEIQLKIY	18	1264	1.7
	6	VLAIEIRLLRQRVIQV	16	1159	1.6
	7	VLKIEIVLEI	10	1021	1.4
	8	IIKIKIELRLIRQQHQIL	18	845	1.2
	9	VIAVRIEIRQLRLQVRQL	18	625	0.9
	10	VEKIEIQVRRVRIKILEV	18	623	0.9
	11	IILVQQRIIVKIQLQLKAV	19	542	0.7
	12	IAIIELELQQVRLKVKAI	18	518	0.7
	13	IIIVQAEIIKIKLQLQKL	18	517	0.7
	14	IIRIEIQLQAKIARH	15	515	0.7
	15	IQELEIRLAIIRQEIQII	18	487	0.7
	16	IKIIQLRVIRLEIQAV	16	470	0.6
	17	IIKIKLRLAEWQLKIIRV	18	468	0.6
	18	IIKLRIKEIEIQIIQV	16	457	0.6
	19	VLKIQLQLKIIQLQVEQL	18	444	0.6
	20	IIKVQLRIEIVQKQVARL	18	434	0.6

Appendix Table 5.3 continued: Top 20 peptide sequences from deep sequencing of the 2.5h library CIS display outputs. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 2.5h library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides. Values are given to 2 s.f. except observation frequencies (raw number).

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
High stringency selection, round R3	1	VLRVRIKIKRELLIIRQKVIQV	22	598	2.1
	2	IIEVQLEIKIVRKQVILL	18	449	1.6
	3	LQLLRLEIKLIQKKVQIVRLQVKQV	25	421	1.5
	4	VIIVQLRLIIVQKRVKLI	18	320	1.1
	5	VLLVEIRWRRIQIQVLRI	18	316	1.1
	6	IIKYEIQLKLVQIEVIRV	18	303	1.1
	7	IIEIEIIIRIKKEIRLI	17	254	0.9
	8	VIIVQRKIQIIQRQVLLV	18	251	0.9
	9	IIKLKIRLQALELQLEVRAL	20	237	0.8
	10	IIIIKLQVKIIRLQLIQI	18	233	0.8
	11	LIEVKLRVQIVRRIKLQVLLV	21	229	0.8
	12	IIKLEIKLRVEIV	13	228	0.8
	13	IIEIELKLRIRLVRKQLLRL	20	220	0.8
	14	VQKLRARIIIIKLKIIQV	18	218	0.8
	15	LRIVQLKIIIIQKEVKRW	18	211	0.7
	16	VIEIEIKLKLVRIKVEQL	18	189	0.7
	17	FIEIEIKVKIVRKEVKEV	18	178	0.6
	18	FIEVRLQLRIQLVRKKVAQV	20	177	0.6
	19	IIQLELRIQIVRKQLKII	18	174	0.6
	20	IIEIRLRVQARIQLIRV	17	172	0.6
High stringency selection, round R4	1	LQLLRLEIKLIQKKVQIVRLQVKQV	25	2446	2.7
	2	VLRVRIKIKRELLIIRQKVIQV	22	2327	2.5
	3	IIEVQLEIKIVRKQVILL	18	1913	2.1
	4	VLLVEIRWRRIQIQVLRI	18	1496	1.6
	5	VIIVQRKIQIIQRQVLLV	18	1394	1.5
	6	VIIVQLRLIIVQKRVKLI	18	1353	1.5
	7	IIEIEIIIRIKKEIRLI	17	907	1.0
	8	IIKLEIKLRVEIV	13	857	0.9
	9	IIKLKIRLQALELQLEVRAL	20	783	0.9
	10	IIKYEIQLKLVQIEVIRV	18	753	0.8
	11	IIEIELKLRIRLVRKQLLRL	20	739	0.8
	12	IIIIKLQVKIIRLQLIQI	18	702	0.8
	13	LIEVKLRVQIVRRIKLQVLLV	21	651	0.7
	14	VQKLRARIIIIKLKIIQV	18	569	0.6
	15	IIQVRIEIQIVKIEVLAL	18	540	0.6
	16	LRIVQLKIIIIQKEVKRW	18	540	0.6
	17	VLIVELRIKLVQLRVQAIQKRVRRV	25	513	0.6
	18	VKLIKIEIIIIQQQVLQV	18	502	0.6
	19	IIIIELKLKQLRIEIREV	18	493	0.5
	20	IIEIQLEYVARVRIEIKEH	19	476	0.5

Appendix Table 5.4: Top 20 peptide sequences from deep sequencing of the 3.4hFosW library CIS display outputs. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 3.4hFosW library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides. Values are given to 2 s.f. except observation frequencies (raw number). **Table continues overleaf.**

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
Low stringency selection, round R3	1	LTALIEQNAALLETLKTLSTEIEKL	30	10892	13
	2	LEALIKSNATLITELKELSKELEKL	30	3997	4.7
	3	LEELIKSNTELLTTLTALSTALEKL	30	3354	3.9
	4	VEALLTTVEELEKSNETLLKVLAELSEEVEKL	37	2719	3.2
	5	VEALLKSNEALIESVKALATELEKL	30	2550	3.0
	6	IEKLIKDNETLLESLKALSKELEKL	30	2343	2.7
	7	LTALIEQNTALLELLKALSTELEKL	30	2216	2.6
	8	LTELVKSNEKLLAELKELSKKLEKL	30	2205	2.6
	9	LEELIASNTELLATLTELSAKLEKL	30	2160	2.5
	10	LTALDASLEELQTSHTTLLEALAALKTELEKL	37	1968	2.3
	11	LEELITANAELLKMLKTLSTTLEKL	30	1807	2.1
	12	IEELITSNATLLTTLKALSEALEKL	30	1653	1.9
	13	LEELIEQNTALCKLLKELSEALEKL	30	1596	1.9
	14	LEELIKANTKLLADLKTLSTELEKL	30	1562	1.8
	15	HETLDASLEELQTSHTTLLEALAALKTELEKL	37	1537	1.8
	16	LEALIAQNAELLKLLKELSTKLEKL	30	1428	1.7
	17	IKKLDTALTTLIEDNAALLEQIKKLSEELEKL	37	1359	1.6
	18	LETLIENNAALLKTLKALAEEIEKL	30	1131	1.3
	19	LEKLITSNTELIKNLKTLSEELEKL	30	1057	1.2
	20	LEELIKFNTELSKQLTALVKALEKL	30	1024	1.2
Low stringency selection, round R4	1	LTALIEQNAALLETLKTLSTEIEKL	30	28198	20
	2	LEALIKSNATLITELKELSKELEKL	30	11142	7.7
	3	LEELIKSNTELLTTLTALSTALEKL	30	5787	4.0
	4	VEALLTTVEELEKSNETLLKVLAELSEEVEKL	37	5284	3.7
	5	LEELIASNTELLATLTELSAKLEKL	30	4969	3.5
	6	LTALDASLEELQTSHTTLLEALAALKTELEKL	37	4672	3.2
	7	IEELITSNATLLTTLKALSEALEKL	30	4175	2.9
	8	HETLDASLEELQTSHTTLLEALAALKTELEKL	37	3900	2.7
	9	LTALIEQNTALLELLKALSTELEKL	30	3896	2.7
	10	IEKLIKDNETLLESLKALSKELEKL	30	3512	2.4
	11	LEELITANAELLKMLKTLSTTLEKL	30	3303	2.3
	12	LEELIKANTKLLADLKTLSTELEKL	30	3221	2.2
	13	LTELVKSNEKLLAELKELSKKLEKL	30	3142	2.2
	14	LEELIEQNTALCKLLKELSEALEKL	30	2995	2.1
	15	LETLIENNAALLKTLKALAEEIEKL	30	2478	1.7
	16	VEALLKSNEALIESVKALATELEKL	30	2302	1.6
	17	LEKLITSNTELIKNLKTLSEELEKL	30	1956	1.4
	18	LTELIADNEKLMTLLTELSATLEKL	30	1768	1.2
	19	IKKLDTALTTLIEDNAALLEQIKKLSEELEKL	37	1539	1.1
	20	LEALIAQNAELLKLLKELSTKLEKL	30	1521	1.1

Appendix Table 5.4 continued: Top 20 peptide sequences from deep sequencing of the 3.4hFosW library CIS display outputs. The most enriched peptide sequences (≥50 observations) identified after CIS display selection rounds R3 or R4 for the 3.4hFosW library, ranked in order of observation frequency. Relative observation frequencies are calculated as a percentage of all enriched peptides. Values are given to 2 s.f. except observation frequencies (raw number).

Selection stringency and round	Rank	Sequence	Length (residues)	Observation frequency	Relative observation frequency (%)
High stringency selection, round R3	1	HKALFKMVKKLGATIKTLHTTIEKL	30	2642	12
	2	DAALFAYIEELRATVKTLHTTLEKL	30	2541	11
	3	LKTLLEMLEKLRADVKTLHTTIEKL	30	1917	8.5
	4	LKKLLEYLETLSAHVTTLHATVEKL	30	1654	7.3
	5	LTKLFTFLEELGTPVKTLHATIEKL	30	1642	7.3
	6	LTKLLKMIETLNANVKTLYTTLEKL	30	981	4.3
	7	LEELIESNEELIKEIATLFKTLEKL	30	789	3.5
	8	VPTALFKYLKALHANIKTLVTTIEKL	29	728	3.2
	9	LETLIADNAALLTSLKELSTALEKL	30	626	2.8
	10	LTELIEFNTTLLASLKELSKKLEKL	30	470	2.1
	11	LKKLIEYLAALSANVKTLVATVEKL	30	453	2.0
	12	ITTLLTIKTKLSTTNAKLTTKIEKL	30	450	2.0
	13	EKKLLEYLKKLSANVKTLHATIEKL	30	444	2.0
	14	LKALLTYLETLHTSVKKLHTTLEKL	30	421	1.9
	15	IKTLITIQTTLCTAQATLVTAIEKL	30	388	1.7
	16	IETLSKYLTKLGADVKTLHATIEKL	30	317	1.4
	17	LAKLLKLIKTLHANVATLHATLEKL	30	304	1.3
	18	LTELIEQNKTLIATLKKLSTELEKL	30	288	1.3
	19	LKKLLTFLEELQTPIKTLHATVEKL	30	268	1.2
	20	SSNTALNTTLKKLTTAITKLVKTIEKL	30	265	1.2
High stringency selection, round R4	1	LKKLLEYLETLSAHVTTLHATVEKL	30	10552	15
	2	HKALFKMVKKLGATIKTLHTTIEKL	30	9617	14
	3	LKTLLEMLEKLRADVKTLHTTIEKL	30	9145	13
	4	DAALFAYIEELRATVKTLHTTLEKL	30	8507	12
	5	LTKLFTFLEELGTPVKTLHATIEKL	30	6019	8.7
	6	LEELIESNEELIKEIATLFKTLEKL	30	2506	3.6
	7	LETLIADNAALLTSLKELSTALEKL	30	2385	3.4
	8	LTKLLKMIETLNANVKTLYTTLEKL	30	2230	3.2
	9	DAALFAYIEELRATVKTLHMTLEKL	30	1820	2.6
	10	VPTALFKYLKALHANIKTLVTTIEKL	29	1498	2.2
	11	LKKLIEYLAALSANVKTLVATVEKL	30	1268	1.8
	12	EKKLLEYLKKLSANVKTLHATIEKL	30	1119	1.6
	13	IKTLITIQTTLCTAQATLVTAIEKL	30	1012	1.5
	14	LKALLTYLETLHTSVKKLHTTLEKL	30	760	1.1
	15	LTELIEFNTTLLASLKELSKKLEKL	30	753	1.1
	16	ITTLLTIKTKLSTTNAKLTTKIEKL	30	692	1.0
	17	IETLSKYLTKLGADVKTLHATIEKL	30	664	1.0
	18	LKKLLTFLEELQTPIKTLHATVEKL	30	645	0.9
	19	SSNTALNTTLKKLTTAITKLVKTIEKL	30	633	0.9
	20	LEALIENNTKLLANLKTLSAALEKL	30	493	0.7

Appendix Table 5.5: Overall composition of peptides from deep sequencing of the 12mer, 16mer, 2.5h and 3.4hFosW library CIS display outputs. Percentages of sequence reads of peptides selected by CIS display after selection rounds R3 or R4 that were observed ≥50 times, 2 – 10 times, or only once. The numbers of sequences observed the stated number of times are given in parentheses, and the total number of sequence reads is given in the rightmost column. Percentages given to 2 s.f.. For 12mer and 16mer selections, the percentages of 'enriched' sequences observed ≥50 times and those observed 2 – 10 times is very low relative to the percentages of peptides observed only once, indicating that these selections have not been successful in enriching particular sequences. Conversely, the 3.4hFosW low stringency selection displays much higher percentages of enriched sequences and similar/marginally higher percentages of sequences observed 2 – 10 times relative to the percentages of peptides observed once, indicating this selection stringency has been successful. Interestingly, 2.5h low and high stringency selections have high percentages of peptides observed ≥50 times and 2 – 10 times relative to peptides observed only once, indicating peptides have been successfully enriched during selection. This conclusion is in contrast to analysis of observation frequencies of peptides observed ≥50 times (Figure 5.5A-D) alone, which indicated poor separation of the most frequently observed peptides from those of lower frequencies, and low frequencies compared to other library selections (i.e. 3.4hFosW in Figure 5.5E-H).

Library Selection	Selection stringency and round	Sequences observed ≥50 times: percentage (number)	Sequences observed 2 – 10 times: percentage (number)	Sequences observed once: percentage (number)	Total number of sequences
12mer	Low stringency, round R3	0.01 (22)	12 (29,442)	88 (219,249)	249,647
	Low stringency, round R4	0.07 (135)	14 (28,869)	85 (173,706)	204,329
	High stringency, round R3	0.00 (10)	6.1 (18,220)	94 (281,713)	300,381
	High stringency, round R4	0.13 (411)	12 (37,340)	87 (282,688)	323,708
16mer	Low stringency, round R3	0.03 (86)	13 (39,274)	87 (263,460)	304,612
	Low stringency, round R4	0.10 (202)	15 (29,820)	84 (171,779)	203,850
	High stringency, round R3	0.01 (54)	12 (43,828)	88 (326,474)	371,531
	High stringency, round R4	0.05 (107)	14 (28,441)	85 (175,634)	205,578
2.5h	Low stringency, round R3	0.32 (223)	31 (21,706)	66 (46,243)	69,884
	Low stringency, round R4	1.0 (532)	28 (14,348)	68 (35,107)	51,981
	High stringency, round R3	0.50 (306)	25 (14,859)	72 (43,354)	60,640
	High stringency, round R4	1.2 (630)	22 (11,585)	73 (38,088)	52,494
3.4hFosW	Low stringency, round R3	0.33 (191)	17 (10,092)	82 (48,190)	58,756
	Low stringency, round R4	0.42 (188)	15 (6590)	84 (37,332)	44,392
	High stringency, round R3	0.09 (72)	21 (17,000)	79 (64,208)	119,434
	High stringency, round R4	0.06 (71)	28 (32,910)	72 (47,980)	81,472

Appendix Table 5.6: Peptides identified in Single Step selection bacterial colonies for initiation of the correct-length PCA for the 2.5h library. Correct-length 18mer peptide sequences observed in sequencing of the 2.5h correct-length PCA Single Step selection colonies that were taken forward into the corresponding Competition Selection are displayed. Rankings from CIS display selection outputs and observation frequency relative to the most frequently observed peptide are also displayed. Three sequences were not observed in the low stringency CIS display output from which these peptides originated, however sequencing likely did not cover the entirety of sequences observed only once. Colonies are numbered very loosely based on speed of colony formation, such that assumptions as to peptide affinity ranking were not considered accurate.

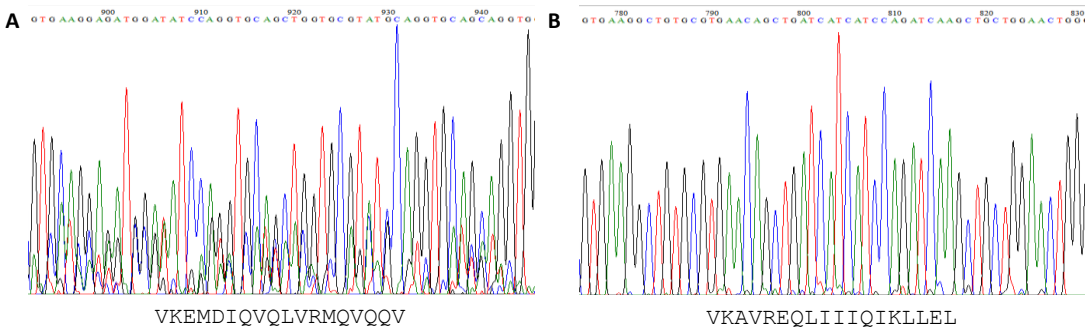
Colony #	Sequence	Ranking in CIS output, observations/observations of #1 peptide
1	VEIVRIEIIQVRQRLLEV	#465, 56/1645
4	VKAVREQLIIIQIKLLEL	#1, 1645/1645
5	VLLLKLEIREVQIEIAAV	#326, 76/1645
8	VLEVRAENIEIQIKIIQL	#154, 126/1645
9	VIKLKIEVQIVRKQIAQV	#430, 60/1645
11	IKRIKLRIQLVQLEVRQI	Not observed
12	LQLVRIRVLQIQLRVLKL	Not observed
13	VQEIEIRVQLVRIQVQLV	#22, 430/1645
15	VIAIEIQLRIVRQQVIKV	#166, 120/1645
17	VIAVRIEIRLLRLQLRQL	#10, 625/1645
18	IKQIELKLRIIQLRVQKV	Not observed
20	VEKVQIRIQLLQLQVIAL	#4, 1350/1645
21	IIQVRLKVIQLRRQLKRV	#28, 367/1645
24	VLQVRIKVKEWQLRIEKV	#187, 113/1645
25	VQIVKAQIIQVQIELRAL	#36, 293/1645
26	VQLVQAQIIKVRLQLQKV	#196, 109/1645
28	VQLVRIEIQELRLEVKAL	#47, 256/1645

Appendix Table 5.7: Peptides identified in Single Step selection bacterial colonies in an attempt to initiate the correct-length PCA for the 3.5h library. All peptide sequences observed in sequencing of the 3.5h correct-length PCA Single Step selection colonies are displayed. Rankings from CIS display selection outputs and observation frequency relative to the most frequently observed peptide are also displayed. Six sequences were not observed in the low stringency CIS display output from which these peptides originated, however sequencing likely did not cover the entirety of sequences present only once in the CIS display output. Colonies are numbered very loosely based on speed of colony formation, such that assumptions as to peptide affinity ranking were not considered accurate.

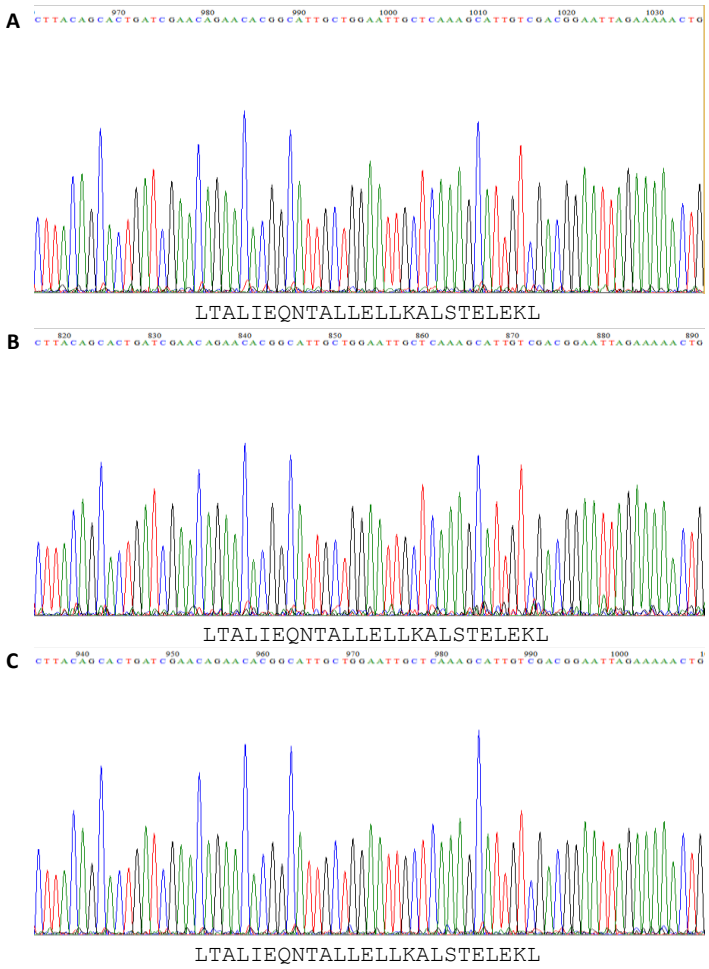
Colony #	Sequence	Ranking in CIS output, observations/observations of #1 peptide
1	IEIIKQQIIIVQARIRAIQLQVKLVRKQVLQV	Not observed
2	MGQVEGEEM*LEIH*EEIR*QIIELEKQIKQI	Mixed read
3	VQEIELQLEELEKQNKLIQEQLIAVQKEIEQV	Position #1, 22,422/22,422
4	VQEIELQLEELEKQNKLIQEQLIAVQKEIEQV	Position #1, 22,422/22,422
5	VLAIEKELEEIELRNAALQLQIQEWQEQIAKLQKLQRL	Position #5, 6747/22,422
6	IKLVQREIAAWRIKNQQLRIEIRQLEAI	Position #2, 13,207/22,422
7	VAKLEIELEIIEIQLQLIREQLKELQKQIQAV	Position #19, 957/22,422
8	IKRLEVRIIEKRIQIVRIQVRIV	Not observed
9	IKKIELQLLEKRIQIVRLKIQI	Not observed
10	IEIIKQQIIIVQARIRAIQLQVKLVRKQVRRV	Not observed
11	VQEIELQLEELEKQNKLIQEQLIAVQKEIEQV	Position #1, 22,422/22,422
12	IEIIKQQIIIVQARIRAIQLQVKLVRKQVLQV	Not observed
13	IKLVQREIAAWRIKNQQLRIEIRQLEAI	Position #2, 13,207/22,422
14	VAAIELEVQELELQNRIIRAQLAIVQAEVEKL	Position #4, 8624/22,422
15	VQEIELQLEELEKQNKLIQEQLIAVQKEIEQV	Position #1, 22,422/22,422
16	IAKLEKRIIELQIKLLQAQVEKVQLEIKAVEEQVRLV	Position #6, 3939/22,422
17	IIRVELQVRIIRRQNALVQRRIRKV	Not observed

Appendix Table 5.8: Homomeric and heteromeric sample helicities and coiled coil formation for CIS display and CIS→PCA selected peptides from the 2.5h, 3.5h and 3.4hFosW libraries. Fractional helicities and 222:208 ratios ($\Theta_{222}/_{208}$, where Θ is ellipticity in mdeg) for peptides alone or an equimolar mixture with cJun as determined by CD spectroscopy at 20 °C. Fractional helicities were calculated according to Equation 4 (see 2.9.6.1). 222:208 ratios are indicative of coiled coil formation where $\Theta_{222}/_{208}>\approx 0.9$ (Zhou *et al.*, 1992, Lau *et al.*, 1984). FosW, 3.4HCIS, 3.4LCIS and 3.4W determined in KPP buffer (10 mM KH_2PO_4/K_2HPO_4 + 100 mM KF), all other peptides determined in "low salt" buffer (10 mM KH_2PO_4/K_2HPO_4). CIS1, CIS21 and CPW peptides from Chapter 4 are provided for comparison. Values are given to 2 s.f.. 2.5pep1, 3.5pep22, 2.4HCIS and 3.4W indicate formation of homodimeric coiled coils (CCs), whilst only 3.4W is suggested to form heterodimeric CCs with cJun by 222:208 ratio. Helicity of heteromeric samples is generally lower than for homomeric samples. CIS display and CIS→PCA selected peptides from the 2.5h, 3.5h and 3.4hFosW libraries display lesser likelihood to form CCs with cJun than Chapter 4 peptides despite appreciable homodimerisation capacity.

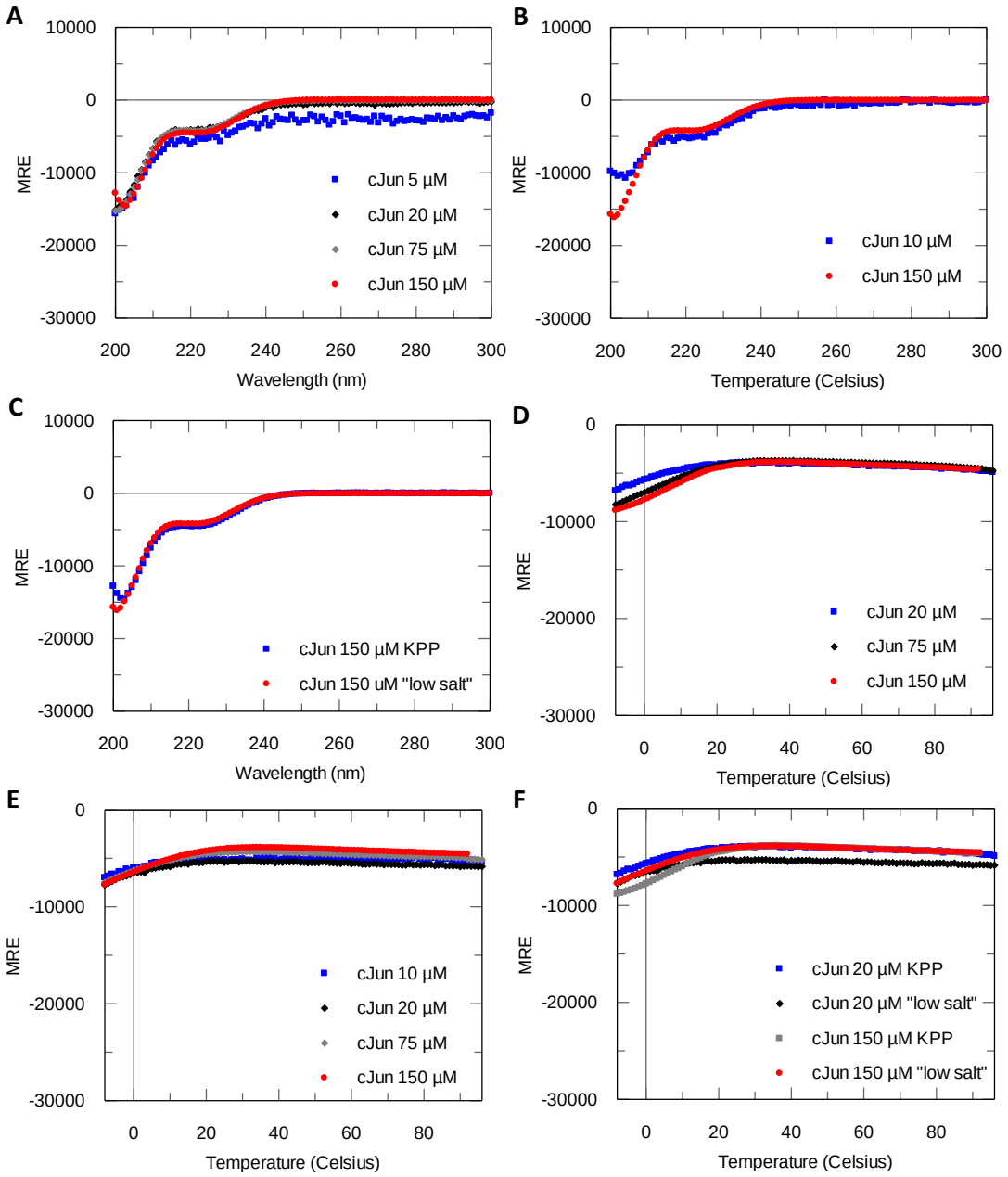
Peptide	**Peptide Fractional Helicity (%) (CD)**	**Peptide $\Theta_{222}/_{208}$ (20°C) (CD)**	**Peptide–cJun Fractional Helicity (%) (CD)**	**Peptide–cJun $\Theta_{222}/_{208}$ (20°C) (CD)**
FosW	41	1.0	37	1.0
CIS1	31	0.69	31	0.71
CIS21	56	0.97	48	0.92
CPW	62	1.0	53	0.95
2.5pep1	5.9	1.2	10	0.47
3.5pep22	7.2	1.2	5.4	0.57
3.5pep23	28	0.85	18	0.70
3.4HCIS	37	0.99	24	0.76
3.4LCIS	50	0.85	38	0.83
3.4W	65	0.93	44	0.87



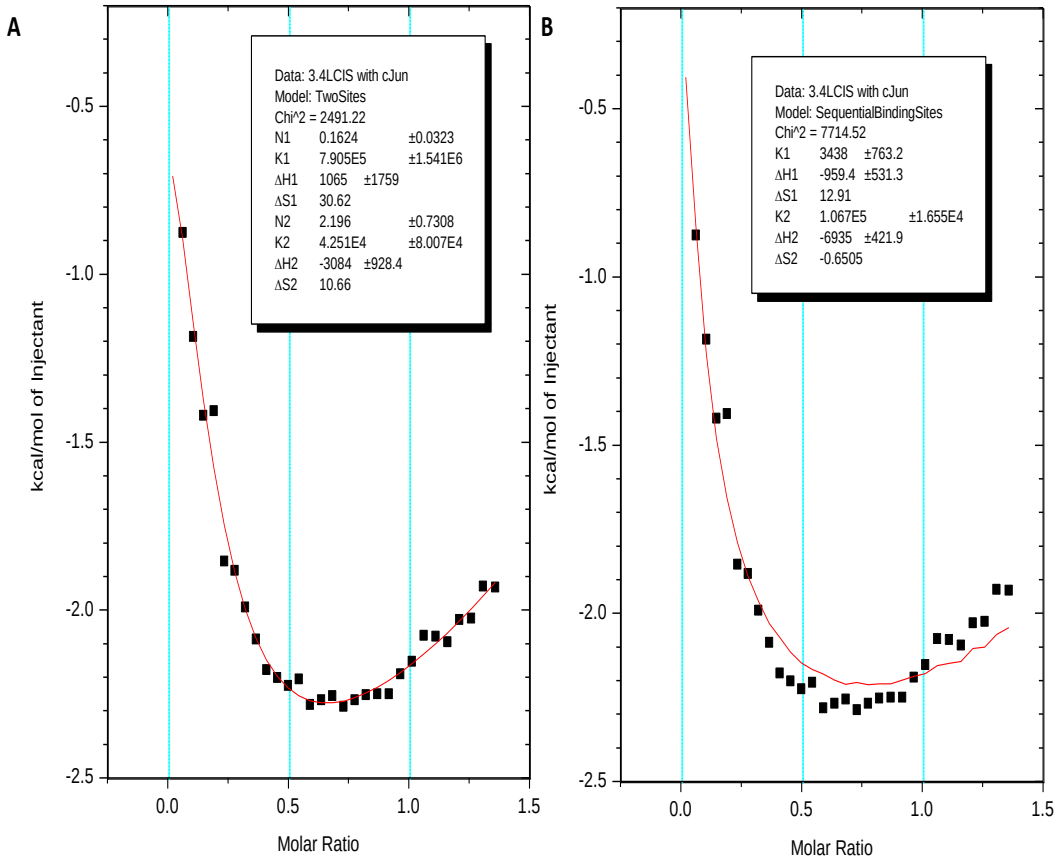
Appendix Figure 5.1: Peptide 2.5pep1 was the winner of correct-length PCA following CIS display low stringency selection. PCA P1 (**A**) and P3 (**B**) sequencing chromatograms. The P1 passage is a mixed read, as expected for pooling of multiple bacterial colonies containing different peptides and subjecting to a single round of Competition Selection, but by passage P3, 2.5pep1 was identified with a clean Sanger sequencing read. Relative fluorescence intensities are shown per base sequenced.



Appendix Figure 5.2: Peptide "3.4Winner" (3.4W) was the clear winner of PCA following CIS display low stringency selection. PCA passage sequencing chromatograms. **A** P1 pool. **B** P2 pool. **C** P3 pool. Relative fluorescence intensities are shown per base sequenced. Throughout P1 – P3, 3.4W dominates the pool.



Appendix Figure 5.3: MREs of cJun assayed by CD at different concentrations over the experimental range reported are equivalent. A cJun CD scans determined at indicated concentrations in KPP buffer. **B** cJun CD scans determined at indicated concentrations in "low salt" buffer. **C** cJun CD scans at 150 µM determined in KPP or "low salt" buffer. **D** cJun CD melts determined at indicated concentrations in KPP buffer. **E** cJun CD melts determined at indicated concentrations in "low salt" buffer. **F** cJun CD scans at indicated concentrations determined in KPP or "low salt" buffer. Mean residue ellipticity (calculated using Equation 3, see 2.9.6) for cJun is virtually identical regardless of concentration or buffer used, allowing comparison across peptides assayed in different buffers and at different concentrations.



Appendix Figure 5.4: ITC data for cJun titrated into 3.4LCIS potentially fits to a two site binding model for a 3-state transition. A Fitting of 3.4LCIS isothermal titration calorimetry (ITC) data for binding to cJun ("Phase 1" of the titration only, see Figure 5.7 in the main text for explanation) to a binding model featuring two distinct, independently-bound binding sites. The unacceptably large errors in K (association constant) in particular do not favour this model despite the closeness of data to the fit line. **B** Fitting of the same ITC data to a binding model featuring binding of one molecule to one site to facilitate binding of the second molecule at the second binding site. The acceptable errors in all parameters and reasonable closeness of the fit line to the data indicates the potential applicability of the sequential binding model to the data rather than a one site binding model for a dimeric interaction which only fits "Phase 2" of the titration and not "Phase 1".